The econometrics of discrete structural change and its applications in models of price overcharge

by

Rossouw van Jaarsveld

Dissertation presented for the degree of Doctor in Philosophy (Economics) in the Faculty of Economics and management sciences at Stellenbosch University

Supervisor: Prof. Willem Hendrik Boshoff

March 2021

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 2020/09/25

Abstract

The econometrics of discrete structural change and its applications in models of price overcharge

R. van Jaarsveld

Department of Economics,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Dissertation: PhD (Econ)

March 2021

This dissertation studies unexplored time series issues related to tests of structural change and the modelling thereof. The practical application is centred around the estimation of price effects in collusive markets. The results, however, is more generally relevant, since structural change is inherent in many economic relationships. The dissertation is organized into three main chapters. The first part deals with the dating of structural breaks, i.e. determining the dates of structural breaks. The second part deals with the impact of non-stationarity when modelling structural changes. The third part deals with a practical application and provides guidance for practitioners.

In cases where the date of the structural change is unknown, two econometric approaches can be followed. One method, is to apply a regime-switching or time varying-parameter model. The second approach is to use structural break tests to determine the break dates, and subsequently construct a dummy variable to control for the breaks in a regression model. In the first part, we investigate how various structural break tests can translate into parameter bias. The results are contrasted to regime-switching and time varying-parameter models to better understand the conditions under which these two approaches can be used interchangeably. Given that structural change may take various forms, we investigate the comparative performance of multiple techniques on different forms of structural change. Specifically, we evaluate various forms of changes in the mean of the data generating process. To assess the performance of these methods, we rely on simulation evidence. Related, we also discuss complications of numerical optimization techniques and regime-switching models in simulation studies. The results are discussed in the framework of overcharge, where parameter estimates play a central role in the punishment of cartels.

The second part deals with the impact of nonstationarity on dummy variable coefficients. This part provides evidence for the conditions under which the distribution of the dummy variable parameter will differ significantly from the t-distribution. This is important since time series models are used in civil litigations by plaintiffs who are required to prove that the cartel had a significant impact on prices. Congruent with the first part, we illustrate how incorrect conclusions can be reached when the dummy variable is misspecified and t-statistic are used to draw inference. In other words, when the break dates are misdated and the nonstationary nature of the data caused the distribution of the dummy variable parameter to differ from the t-distribution. Related, we show how cointegration tests are influenced when the break dates are misspecified. We provide a discussion of the extent to which error correction modelling can be used to address some of these issues and emphasize specification problems that are specific to the estimation of overcharge.

To demonstrate the practical insights of the simulation studies in the first and second part, I apply the various techniques to a European competition case. This part also provides guidance for practitioners.

Opsomming

The econometrics of discrete structural change and its applications in models of price overcharge

("Die ekonometrie van diskrete strukturele verandering en toepassings in modelle van prysoorverhalings")

R. van Jaarsveld

Department of Economics,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Proefskrif: PhD (Econ)

Maart 2021

Die proefskrif bestudeer probleme in tydreeks analise wat met toetse vir strukturele verandering, en die modellering daarvan, verband hou. Die praktiese toepassing fokus op die beraming van prysoorverhaling in markte waar mededingers saamspan. Buiten hierdie praktiese fokus, is die resultate van algemene belang, aangesien 'n verskeidenheid van ekonomiese verwantskappe dikwels deur struktuurveranderinge geraak word. Die proefskrif bestaan uit drie dele. Die eerste deel handel oor die datering van strukturele breke, met ander woorde die bepaling van die datums waarop die strukturele verandering neerslag in ekonomiese verwantskappe vind. Die tweede deel handel oor die effekte van nie stationêr in modelle van strukturele verandering. Die derde deel bespreek 'n praktiese toepassing.

In gevalle waar die datums van die breekpunte onbekend is, kan een van twee ekonometriese benaderings gevolg word. Een metode is om 'n regimewisselling model of tydwisselende parameters model te gebruik. 'n Tweede metode is om 'n strukturelebreek toets uit te voer ten einde die breekdatums vas te stel en dan 'n fopveranderlike, o.g.v. die resultate van hierdie datums, te skep. Hierdie fopveranderlike kan dan gebruik word om vir die strukturele breuke in 'n regressiemodel te kontroleer. In die eerste deel van die proefskrif evalueer ons hoe verskeie strukturele breekpunt toetse parameter sydigheid kan veroorsaak. Die resultate word met regimewisselling en tydwisselende parameter modelle vergelyk ten einde te bepaal onder watter omstandighede die

verskillende metodes uitruilbaar is. Gegewe dat strukturele verandering verskeie vorme kan aanneem, word alternatiewe tegnieke vir verskillende vorme van strukturele verandering ondersoek. Die soeklig val op verskeie gevalle waar die verandering in die gemiddelde van die datavoortbrengende proses is. Om die gedrag van verskeie metodes te evalueer word daar van simulasie gebruik gemaak. Die proefskrif bevat ook 'n bespreking van verwante probleme in verband met numeriese optimering en regimewisselling modelle in simulasie studies. Die resultate word bespreek in die konteks van prys oorverhalings waar parameter skattings 'n belangrike rol speel in die vergelding van kartels.

Die tweede deel handel oor die impak van nie stationêr op die koëffisiënte van fopveranderlikes in regressiemodelle. Hierdie deel omvat bewyse vir die omstandighede waaronder die verdeling van die fopveranderlike koëffisiënt aansienlik van 'n t-verdeling verskil. Dit is belangrik aangesien tydreeksmodelle gebruik word deur klaers in siviele litigasie om bewyse te lewer dat samespanning 'n betekenisvolle impak op pryse gehad het. Ooreenkomstig met die eerste deel, wys hierdie deel hoe foutiewe gevolgtrekkings kan volg wanneer die fopveranderlike verkeerd gespesifiseer word, dit wil sê, wanneer die breekpunte verkeerd gedateer word en die niestationêre eienskappe van die tydreeks veroorsaak dat die fopveranderlike se parameterverdeling verskil van 'n t-verdeling. Die verwante probleem van hoe koïntegrasie toetse beïnvloed word wanneer die breekpuntdatums verkeerd bepaal is, word ook ondersoek. Die proefskrif bespreek voorts hoe foutkorreksie modelle gebruik kan word om van die hierdie probleme aan te spreek en beklemtoon spesifikasieprobleme wat spesifiek tot die beraming van prysoorverhaling is.

Om die praktiese insig van die simulasie studies in die eerste en tweede deel te illustreer pas ek die verskeie metodes toe op 'n Europese kompetisie geval. In die deel verskaf ek ook 'n raamwerk vir praktisyns.

Acknowledgements

This PhD has been a rough journey full of sleepless nights, self-doubt and new forms of anxiety. I did not win this battle on my own and would like to express my gratitude to everyone who helped me along the way. First, I need to thank my supervisor, Willem Boshoff, without whom this PhD would never have started in the first place. Thank you for believing in me when I certainly did not, for always being patient and kind. Your understanding when I suffered various medical setbacks meant a lot. The advice and guidance outside of this PhD will never be forgotten and I will always be grateful.

My partner Lauren Kuhn deserves massive praise. Thank you for giving me a pass on household chores and never being upset when my alarm went off at horrendous hours of the morning. Thank you for being my emotional support when this process had me beaten down at times. I sincerely appreciate every stimulating discussion and debate about econometrics we had over the past years. My gratitude is also extended to the Kuhn family who has been immensely supportive.

I would also like to thank my parents without whom I never would have been able to pursue further studies. Thank you for always being willing to create an environment that allowed me to focus on my studies and your support throughout the years. I also want to thank my brother and sister who never stopped asking when I will finally be done. I am done now. I think.

My friends at the CCLE also have my appreciation. Thank you for the support and always being willing to engage with me on various subjects in the field. Thank you to the participants of the weekly CCLE seminars who gave valuable feedback in the earlier stages of this research. I am grateful for my friends in the 2017 GEM PhD cohort. Jaco Franken should be applauded for the immensely supportive role he plays in this group.

Lastly, I would like to thank Daan Steenkamp who was more than willing to delay some of our projects and give me extended leave. I would also like to thank Daan for being a good friend and mentor during this time.

Dedications

To Lauren and my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A standard assumption in regression analysis is that the model parameters are constant over the entire sample period. However, the relationships represented by regression models can, and often do, change over time. Wars, institutional changes, price shocks, technological developments and monetary policy changes are but a few examples. These types of changes will cause the data to have different characteristics for some or all subsamples. This dissertation deals with the detection and modelling of structural change.

In regression models, different subsample data characteristics will cause parameter instability. If ignored, the model's functional shape will be incorrect, and the resulting estimates can be severely biased, causing inferential distortions. Thus, when estimating regressions in the presence of structural breaks it is important to consider differentiating of the slope parameters and constant term for subsample periods that are different from one another.

There is an abundance of empirical evidence indicating that many economic time series relationships are subject to structural shifts (see, for example, Stock and Watson, 1996; McConnell and Perez-Quiros, 2000; Sensier and Dijk, 2004; Bataa *et al.*, 2013). Since the seminal work of Page (1955, 1957), there have been extensive developments in the statistical and econometric literature to uncover the presence of structural change. However, when modelling economic relationships, it is not enough to know only whether a structural break is present or not. In order to adequately control for structural breaks, it is pivotal to know the timing thereof. Tests to determine the timing of structural breaks date back to the pioneering work of Chow (1960), Quandt (1958, 1960), Gardner (1969) and Brown *et al.* (1975). This gave rise to the development of several tests for structural breaks. These include Bai-Perron break test (Bai and Perron, 1998), cumulative sum (Ploberger and Krämer, 1992), moving sums (Chu *et al.*, 1995), and sequential F-tests (Andrews and Ploberger, 1994) (which will receive further attention in this dissertation), as well as other methods such as those based on quantile functions Qu (2008).

Once the break dates are known, there are two ways to deal with structural change in regression analysis. One method is to assume that structural shifts

occur at discrete times or, alternatively, to assume that change occurs for all time periods. The latter case involves the use of time-varying-parameter (TVP) models (see for example, Tucci, 1995). If changes are assumed to be discrete, they can be modelled either in a deterministic or stochastic manner. If changes are deterministic, the shifts can be modelled by using a dummy variable. In the stochastic case, changes are modelled through a transitional-probability matrix for a specified finite number of regimes. This will ensure that the model moves from one regime to another in a stochastic manner.

In practice, there is a two-step modelling procedure to deal with discrete structural breaks at unknown dates. The first step is to determine the break dates by using a structural break test. The second step is to construct a dummy variable that is informed by the structural-break-test results. The dummy variable can then be used in a regression model to control for the structural change. In Chapter 2, I consider how structural-break-test results map onto the parameter estimates of dummy variables when the dummy variable was constructed by relying on the test results. This is a departure from the standard literature that investigates the performance of statistical tests by calculating the type I and II error rates under various conditions. The question is, therefore, How structural break-tests translate into parameter bias in subsequent regressions?

An alternative modelling approach to dealing with structural breaks at unknown dates, is to estimate a regime-switching model. In this approach, the model will control for the effects of the structural change while simultaneously estimating the timing thereof. An unanswered question is; How would using a regime-switching model compare with the previously mentioned dummy variable approach? This question is dealt with also in Chapter 2.

When dealing with structural breaks, special consideration needs to be given to the presence of unit roots. If a time series follows a unit root process, the series integrates its previous value over time. As a result, any shock will accumulate over time and have permanent effects on future values. Due to this feature, unit root processes can easily be mistaken as a process that contains structural breaks. Conversely, structural breaks also influence unit root tests. Most unit root tests will favour the hypothesis of a unit root when the true process contains a structural break and is not a unit root process. It is therefore important to consider the interplay between unit roots and structural change.

Given that dummy variables are often used to account for structural change, it is important to consider how dummy variable parameters are influenced by unit roots. There are two dimensions to consider in this regard. First, since the seminal work of Park and Phillips (1988, 1989), the asymptotic properties of regression coefficients in the presence of unit roots has been well understood. However, not much attention has been given to the associated effects on dummy variable coefficients. Chapter 3 deals with this oversight. The second dimension to consider is how structural breaks influence cointegration testing. These tests are standard procedure when regression variables contain unit roots to

ensure that the resulting estimates are not spurious. Structural breaks can have an impact on the size and power of cointegration tests (Campos *et al.*, 1996). An area where further research is required is how the specification of dummy variables influences cointegration test results. This question is also addressed in Chapter 3.

A field of application where the modelling of structural breaks is keenly relied on is the punishment of cartels. This dissertation focuses on this particular field. In section 1.1, I explain how structural breaks feature in this field. In sections 1.2 and 1.3 I explain how the above-mentioned questions related to chapter 2 and 3 are salient issues in the punishment of cartels. In section 1.5, I provide an outline for the remainder of this dissertation.

1.1 Cartel damages

In markets with low levels of competition competing firms can exert market power and collectively increase their prices. Firms can obtain this situation by, for example, forming a price-fixing cartel whereby they agree to designated price levels. As a result of increased prices that are not driven by market determinants, anti-trust law presumes such cartel behaviour to be harmful to consumers. Cartels are therefore illegal in most jurisdictions (see for example the Competition acts of 1998 in South Africa and the United Kingdom, Article 81 in the European Union and the Sherman Antitrust Act of 1890 in the United States). Despite anti-trust agencies' active prosecution, cartels continue to form – see Connor (2014), which summarises 2,041 cartel judicial decisions.

Economic theory is traditionally relied upon to motivate the negative effects that cartels have on welfare and to provide the rationale for their deterrence. The total welfare effect of cartels consists of two components: one, a decrease in total welfare created by the market and, two, the redistribution of rent to the firms from the consumers. The welfare effect can be approximated by a monopolist setting, where prices are set at the point where marginal cost equals marginal revenue.

The consequences of a cartel on prices and quantity are illustrated in Figure 1.1. As can be seen in this figure, what is observed in a cartelized market is not only an increase in price but also a reduction in total quantity supplied. This is illustrated in Figure 1.1 by consumers purchasing Q^1 units at price P^1, instead of Q^0 units at P^1 under competition. Hence, the total damage to consumers is the lost consumer surplus, which is represented by the area $A+B$. Area A represents the rent transfer from consumers to producers, while area B is the deadweight loss.

Private enforcement in the form of civil litigations allows cartel customers to seek reparations for the damages they suffered due to the illicit cartel behaviour. When calculating damages, area B is typically ignored, since damages are generally defined as the illegal appropriation of profits. In practice, it is

assumed that the appropriated profits of the firm and the damage suffered by the consumer are equivalent (area A). This is commonly referred to as the overcharge. The overcharge for a single unit is given by $P^1 - P^0$, while the total overcharge is given by $Q^1(P^1 - P^0)$. In simplistic terms, the overcharge is the amount that consumers overpaid as a result of the cartel.

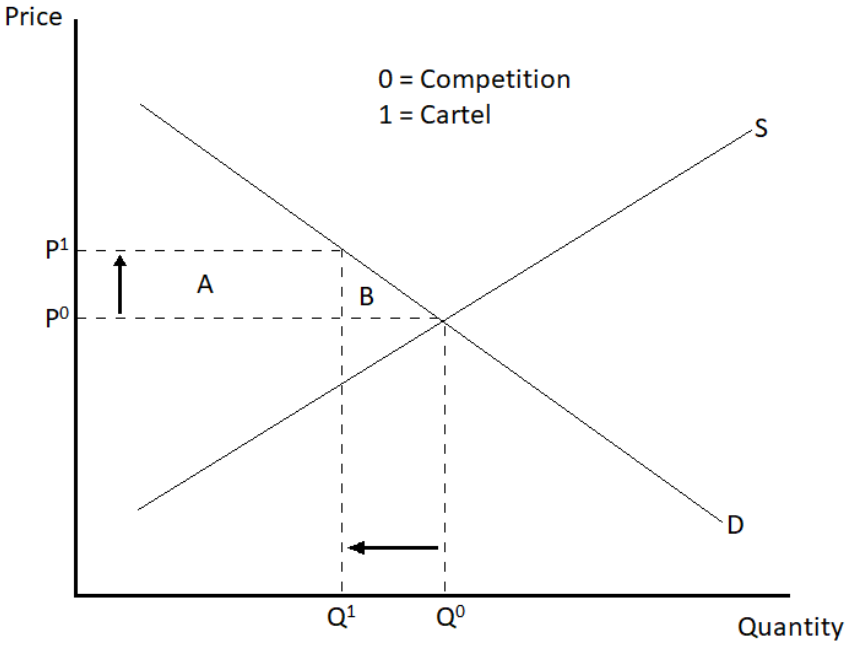


Figure 1.1: Cartel welfare effects

Rubinfeld (2012) points out that to receive reparation in private actions the plaintiff must prove three elements: antitrust violation, antitrust injury, and damages. The focus of this dissertation is on estimating the third element. It is important to note that a universal overcharge does not exist. Therefore, it is estimated on a case-by-case basis. Having robust methods and understanding their shortcomings are important since multi-billion-rand fines are determined based on their results.

In practice, there are several methods used to determine overcharge. The most commonly used make use of *benchmarks* and *yardsticks* (see McCrary and Rubinfeld, 2014). In a typical yardstick approach, data from a comparable market – that is external to the infringement – is used to calculate the price overcharge. In contrast, the benchmark approaches only uses data from the market at issue and compare prices in the contravention period to prices outside of the contravention period. The main focus of this dissertation is on the

benchmark approach, although, a brief discussion of the yardstick approach is included.

A typical yardstick approach would use data from a comparable market to estimate a 'but-for' price. That is, the price that would have prevailed if the contravention had not occurred. The comparable market should, ideally, display the same demand, cost, and competitive behaviour than that would have prevailed in the market in which the illicit behaviour occurred. The product being examined in the comparable market can be the same product that is sold in a different geographical market, or a similar product that is sold within the same geographic market. The drawback of this approach is unambiguous. Products sold in different geographical markets are rarely subject to the same cost, demand, and competition structures. Furthermore, anti-trust authorities do not have legal grounds to demand data from markets outside of the contravention market, and in some cases the required comparable market data does not exist.

With the benchmark approach, data from the same market (such as prices before and/or after the contravention period) is used as a benchmark. Conceptually, the idea is to construct a counter-factual scenario that would have prevailed in the absence of collusion and to compare that to the actual collusive outcome. A simplistic approach would be to compare the average price during the collusive period to the average price before and/or after the collusive period to construct an average overcharge. While this simple approach can provide an initial indication of possible overcharge, it is seldom used in practice since simple price comparisons can be misleading – see, for example Frank and Lademann (2010) and Friederiszick and Röller (2010). Prices are determined by a variety of cost, demand and market structure factors. To appropriately isolate and determine the effect that a cartel had on prices, these factors need to be controlled for. This has made regression analysis the de facto method in estimating of cartel overcharges (McCrary and Rubinfeld, 2014, 63).

The two standard benchmark approaches are the *dummy variable approach* and the *forecasting approach*. Both approaches estimate a reduced-form multivariate regression equation where prices are represented as a function of cost and demand drivers. In practice, the reduced form regression is typically preferred over a structural price equation, since the results are more robust to small changes in the specification, and the number of data requirements are fewer than for a structural equation (Davis and Garcés, 2009, 357).

In the dummy variable approach, a regression model is estimated using data that compromises of the entire sample period – that is the competitive and collusive periods – for which the data is available. The overcharge is obtained by including a dummy variable that spans over the collusive period. Therefore, controlling for various demand and cost shifters, the overcharge is considered to be the mean shift in price over the collusive period.

A general dummy variable approach would involve estimating an equation of the following form:

$$p_t = \alpha + \delta D_t + \boldsymbol{\beta_1'}\mathbf{c_t} + \boldsymbol{\beta_2'}\mathbf{d_t} + \epsilon_t, \tag{1.1}$$

where p_t denotes price, $\mathbf{c_t}$ and $\mathbf{d_t}$ are vectors of cost and demand drivers, α represents the intercept, ϵ_t is the error term, and D_t is a dummy variable equal to 1 in the collusive periods and 0 in the non-collusive periods. The overcharge for a single unit over the collusive period is captured by δ. Typical cost drivers are variables such as labour, raw materials, consumable manufacturing supplies and general overheads. Examples of demand drivers are variables such as sales quantity and production capacity. In cases where the product is used in upstream production, the upstream production capacity or sales quantities can be used as drivers of demand. The approach in practice is not to include all cost and demand drivers but only a parsimonious set of such drivers that account for the constraints imposed by the dataset.

The total overcharge is calculated as:

$$\hat{OC}_1 = \hat{\delta} \sum_{t}^{T} D_t Q_t, \tag{1.2}$$

where $\hat{\delta}$ is the estimated parameter from equation 1.1 and Q_t represents output.

In the forecasting approach, a regression model is estimated using data outside of the contravention period. Subsequently, the estimated model is used where the actual variable outcomes during the collusive period are substituted in the estimated model to construct a counter-factual price that would have prevailed in the absence of collusion.

The standard forecasting approach would estimate the following regression equation using data outside of the contravention period:

$$p_t = \alpha + \boldsymbol{\beta_1'}\mathbf{c_t} + \boldsymbol{\beta_2'}\mathbf{d_t} + \epsilon_t, \tag{1.3}$$

where the variables are the same as in equation 1.1. Once equation 1.3 is estimated using data outside of the collusive period, the parameters in equation 1.3 are used to forecast prices during the collusive period ($\hat{p_t}^f$) – that is the price that would have prevailed in the absence of collusion. In simple terms, the values of the cost and demand drivers during the collusive period are substituted into equation 1.3. Total overcharge is then calculated as:

$$\hat{OC}_2 = \sum_{t}^{T} D_t(p_t - \hat{p_t}^f)Q_t \tag{1.4}$$

As a result of the methodological framework, benchmark methods use time series data. It is therefore, important to account for the various time series properties and assumptions that are required for robust results. The main focus of this dissertation is on the dummy variable approach and the related time series issues.

The focus is on single equation models for a number of reasons. First, guidelines for quantifying harm due to competition law infringements refer exclusively to single equation methods. This focus reflects the position in literature and in practice where single equation methods predominate (for example, the summary of the empirical methods for damage estimation presented by McCrary and Rubinfeld (2014) focus exclusively on single equation methods). Second, macro economic variables are often used as demand drivers in models for estimating overcharge. For example, Boshoff (2015) used aggregate construction activity as a demand driver for bitumen prices and Boshoff and Van Jaarsveld (2019) used a house price index as a driver for cement prices. In these cases, the demand drivers are often less responsive to price developments in particular markets. Last, the empirical results in this dissertation confirm the suitability of single equation methods. A practical example is given in chapter 4. Even so multiple equation methods might well be useful in estimating damages. This matter is left for future research.

It is evident that there is a direct relationship between overcharge estimation and modelling structural breaks. If there are overcharges present in a price series, the implicit implication is that, at some point in time, there was a mean shift in the price level. The mean shift in the price level will therefore cause a structural break in the econometric relationship between a price series and its determinants. Within the time series literature, the dummy variable approach is the same as modelling a structural break in the regression equation.

There are two important econometric issues in estimating overcharge that require further investigation. The first is the dating of the start and end points of overcharge, i.e. determining the dates of structural breaks. The second issue relates to how unit roots can affect the dummy variable approach. Sections 1.2 and 1.3 elaborate on these two issues.

1.2 Cartel dating

In the European Commission's practical guidelines for cartel damage estimation, the Commission states that the start and end dates of the infringement period are not always easily identifiable (European Commission, 2013, 17-18). There are a variety of reasons for this uncertainty. In some cases, the start and end period of the infringement may begin or end gradually. In markets with longer-term contracts, it may take time before the cartel is able to effectively increase its prices. The standard approach is often to take the legal infringement period from documentary evidence as the start and end dates in estimating the overcharge.

In both benchmark approaches, the start and end periods are key elements in the result. It should be recalled that both benchmark approaches essentially compare prices during the competitive period to those during the collusive period. Therefore, if the starting period in the estimation is taken as an

earlier date than the actual effective date on which the cartel raised prices, the inclusion of lower prices in the collusive period will cause underestimation of the overcharge. Likewise, if the cartel maintained higher prices after the legal infringement period and the higher prices are included in the competitive period the result will also be a lower overcharge. It is, therefore, important to empirically establish the start and end periods of the cartel effect, based on statistical tests using the data at hand. This process is referred to as cartel dating.

Within the time series literature, cartel dating is in essence the same as testing for structural breaks. The underlying idea is that cartels arbitrarily increase the mean level of the price at an unknown date, and the breakdown of the cartel will cause the mean level of prices to return to competitive levels. Hence, structural break tests can be used to empirically determine the effective start and end date on which the cartel affected prices.

Overcharge estimation can be thought of as a two-step process. First, cartel dating is performed to determine the start and end periods during which the cartel was effectively manipulating prices. In the second step, an econometric model is estimated using a dummy variable to control for the mean shift in prices or the econometric model is used to forecast the but-for price. This dissertation is mainly focused on the use of dummy variables, although, similar issues will prevail when following the forecasting approach.

If the dummy variables used to control for structural breaks in regression models do not match the effective periods of the breaks, the obvious consequence will be that the resulting parameter estimates will be biased. Structural break tests that determines the break dates inherently carry the risk of misdating the true break dates. If misdating occurred, the overcharge estimation will be biased since the estimation relies on the structural break test results in the second step. It is, therefore, important to understand how statistical tests of dating structural breaks translate into parameter estimates in overcharge estimation. Chapter 2 deals with this issue. Given that structural change may take various forms (and may not involve a hard and sudden shift), I investigated the comparative performance of various techniques, to identify the conditions under which they yield similar results. For this purpose, I relied on simulation evidence.

The construction of simulation models featuring structural breaks is non-trivial, given that the nature of the shift and accompanying effects has a significant impact on the performance of various techniques. The literature typically performs Monte Carlo simulations, testing a specific property of a single method in isolation – see, for example, Bai (1995), Wang (2006), Antoshin *et al.* (2008), Prodan (2008) and Groen *et al.* (2013). Chapter 2 contributes to this literature by comparing various methods and their relative performance in detecting different types of mean shifts in the data generating process (DGP). Additionally, Chapter 2 explains the novel perspective of the influence that structural break test results have on parameter estimates in subsequent re-

gressions.

Related to the dating of structural breaks are regime-switching models. While regime-switching models do not explicitly test for the date of change, these models incorporate structural changes without prior testing of their start and end dates, see for example (Krämer, 2012, 29-38) and (Enders, 2008, 407-474). In Chapter 2, I also make a contribution to this literature by comparing the estimation results of models that rely on structural break tests to the estimation results of regime-switching models. In other words, I compare the results of regressions – where the specification is informed by structural break tests – with the results from regime-switching models. Furthermore, Chapter 2 contains a discussion on the choice of regime-switching model, software, and numerical optimization techniques and how these choices translates to the simulation results.

In chapter 2, four structural break tests and two non-linear models are compared. As motivated in Chapter 2, I considered the following structural break tests: the Bai-Perron break test (BP), cumulative sum (CUSUM), moving sums (MOSUM), and F-tests. For the non-linear models I consider the Markov-switching (MS) and time-varying parameter model. The choice of methods to investigate is motivated by the cartel damages literature and the requirements of practice which is further discussed in chapter 2.

The results in chapter 2 indicate that, in general, the BP and MS approaches will provide the least biased parameter estimates compared to the other methods. In the most basic case, where the structural breaks are driven by deterministic processes, the BP method will provide the most reliable estimates. While the MS model produces good results for all the different types of breaks, there is much more variation in the estimates. This implies that while it is a robust method regardless of the nature of the structural change, greater care needs to be taken when interpreting the results.

Therefore, the contributions can be summarized as follows. First, the parameter estimates vary, implying that misdating of the structural breaks are dependent on the nature of the DGP and type of test employed. Second, when comparing estimates informed by structural break tests to regime-switching models, the two methods do not perform similarly. In cases where there are multiple structural breaks the MS model outperforms structural break tests in terms of bias. When there structural breaks are less frequent, the MS model still performs well but the estimates have significantly higher variance than the BP approach. Last, comparing the methods in a typical horse race, there is no single method that will always provide the most accurate estimates.

1.3 Dummy variables and unit roots

When estimating overcharges using the dummy variable approach, an assumption is made that the data being used is stationary (Boswijk *et al.*, 2019, 8-9).

A common feature of time series is that the data is often nonstationary due to unit roots. Nelson and Plosser (1982) provided statistical evidence that many economic time series contain unit roots. In terms of modelling time series with unit roots, understanding the effects thereof is crucial to applying the correct modelling framework and to interpreting the results. There has been extensive literature on the effects of unit roots and how to deal with them in regressions – see Banerjee (1993) for a comprehensive summary and Juselius (2017) for a summary of more recent developments. Despite the vast literature on the subject, less attention has been paid to the effects that unit roots have on dummy variable coefficients. Chapter 3 aims to address this oversight.

Chapter 3 consider the effects that unit roots have on dummy variable coefficients in cointegrated regressions. It is well known that the long-run parameters in cointegrated regressions do not have standard t-distributions (Park and Phillips, 1988, 1989). The asymptotic distribution of the regression statistics is usually non-normal and depends on nuisance parameters. An unanswered question is How much will the distribution of dummy variable parameters in cointegrated regressions differ from the standard t-distributions? Chapter 3 provides an answer to this question. In chapter 3, I provide empirical evidence for the conditions under which the distribution of the dummy variable parameter will differ significantly from standard t-distributions. Additionally, I demonstrate the conditions under which a t-test will show larger size distortions and lead to incorrect conclusions regarding the significance of dummy variable coefficients.

In the simulation exercise of chapter 3, I vary both the total sample size (T) and length of the structural break (D_t). The results show that for small values of T and (D_t) the 'true' distribution of the dummy variable parameter differs significantly from a Student's t-distribution. The difference between the distributions is also shown to be statistically significant in almost all of the simulations. Consistent with this finding, a large size distortion occurs when using a t-test for small values of T and $T(D_t)$. However, when T and (D_t) is large, the difference between the 'true' distribution of the dummy variable parameter and Student's t-distribution becomes much smaller. This result also holds when evaluating the size distortion of using a t-test when the values of T and $T(D_t)$ are large.

The first contribution is that when dealing with unit roots, special care needs to be taken when drawing inference on the statistical significance of the long-run parameters. This is an important consideration in overcharge estimations because econometric models are used to infer the presence and size of the increased prices. As stated by Rubinfeld (2012), in private damage actions the plaintiff needs to demonstrate that the cartel effectively managed to increase prices. Hence, the statistical significance of the dummy variable parameter is paramount.

It has become standard practice since the seminal work of Granger and Newbold (1974) to perform cointegration tests when the data contains unit

roots . Cointegration testing is an important consideration to ensure that the results are not spurious. The literature has given consideration to cointegration testing in the presence of structural change (Gregory and Hansen, 1996; Gregory *et al.*, 1996). The effects of uncontrolled structural breaks on cointegration tests are well understood. However, the effects of misdating the breaks and then performing cointegration testing requires further understanding. More specifically, I am interested in understanding the effects on cointegration tests when a dummy variable is used to control for a structural break but the dummy variable does not control for the correct periods of the breaks. This is closely related to the issue of misdating breaks in Chapter 2.

The second contribution is that it might seem trivial that cointegration tests would not reject the null hypothesis of no-cointegration when the break dates are incorrectly specified. However, the simulation results indicate otherwise. I find that the bounds test of Pesaran *et al.* (2001) has extremely low power when the misspecification of the dummy variable is large and at the start of the true break period. When the misdating is at the end of the true break period, the bounds test has extremely low power regardless of the extent of misdating. This result highlights the importance of correctly dating structural breaks as discussed in Chapter 2. It emphasizes the importance of using appropriate methods to determine the start and end points of structural breaks, especially when the data contains unit roots. Failure to do so will lead not only to biased estimation of the effects, but also to subsequent specification tests – in this case cointegration tests – possibly falsely indicating that the model is well specified. In the context of overcharge estimation, it is important for practitioners to take note of this result, since biased estimates could be defended on the basis that they pass the required cointegration tests, and that the model is therefore correctly specified.

1.4 Practical application

Given the complications presented in chapter 2 and 3, it is important to understand how the results are relevant for practice. This is illustrated in chapter 4. Using data from the European sodium chlorate cartel, each of the methods in chapter 2 is applied to estimate the overcharge. For this particular application, the data does not appear to be nonstationary. Therefore, chapter 4 does not illustrate the consequences of unit roots on the dummy variable coefficient. However, this chapter provides a discussion on the complexities of dealing with unit roots in this setting.

The results in chapter 4 show that different overcharge estimates will be obtained when depending on the type of structural break test or model. For this case, the BP and MS approaches produces the most sensible estimates. When using the MS model, the estimated overcharge is almost twice the size of the BP result. This is because the MS model incorporates the higher prices

towards the end of the sample as part of the collusive price regime. While defending the inclusion of the later period as part of the collusive price might be difficult, it raises interesting questions on the permanent effects that cartels can have on future prices even after their disbandment.

To illustrate the complexities of dealing with unit roots, this chapter provides a brief discussion of the relevant literature. Due to the nature of cartel price data, a unit root test that allows for multiple structural changes is required. The results show that not adequately controlling for the structural breaks in the unit root test the incorrect conclusion that the data contains a unit root when it does not is often made.

1.5 Outline

In summary, there are two econometric issues that I consider when using the dummy variable approach in overcharge estimation. The first is how various structural break-tests perform when their results are used to inform the start and end periods of the overcharge in subsequent regressions. These results are also compared with regime-switching models that can simultaneously determine the timing of breaks and control for the effects thereof. The second issue is how dummy variable parameters are influenced by the presence of unit roots. Given that in applied work, cointegration testing is required when dealing with unit roots, I also investigate how misspecification of the dummy variable influences cointegration tests.

As mentioned at the beginning of this chapter, there is a strong relationship between these two phenomena. A unit root process can easily be mistaken for a process containing a structural break, and time series that contain structural breaks can easily be mistaken for a unit root process. As a result of this interplay, many of the methods found in the literatures on both of these issues have developed in parallel and rely on the same statistical theories.

While I mainly focused on the applications in overcharge estimation, the results are more generally relevant. Many economic time series are subject to structural breaks and unit roots. Determining the start and end dates of structural breaks, using this information to construct a dummy variable, and handling a dummy variable when the data contains unit roots are pertinent issues in applied work.

The rest of this dissertation is organized as follows: Chapter 2 considers how structural-break-test results translate into overcharge estimations. The chapter provides a general overview of the principles and importance of cartel dating in section 2.2 with a discussion of the relevant literature in section 2.3. In Chapter 2 I consider four types of structural breaks that are motivated by cases in the overcharge literature. Each of these cases and their motivations is discussed in section 2.2. Section 2.4 explains the simulation set-up, with special attention given to how I simulate each of the four different types of structural

breaks. In section 2.5 I report the results of the simulations and compare the performance of various structural break tests and regime-switching models when used to estimate overcharges. Section 2.6 explains complications that are encountered during simulation studies when dealing with regime-switching models which are reliant on numerical optimization algorithms. Some of these complications explain the simulation results, and contain valuable information on how these models will perform in practice. Section 2.7 summarizes the findings of Chapter 2.

Chapter 3 deals with the impact of unit roots on the distribution of dummy variable coefficients. Related to this, the chapter also studies to what extent cointegration tests are influenced by the specification of the dummy variable. Section 3.3 provides a technical overview of how parameter distributions are influenced by unit roots. This is followed by a short explanation in section 3.4 on why parameter distributions are particularly important in overcharge estimations. Section 3.2 discusses the related literature and how chapter 3 contributes to existing work. In section 3.5 I explain the simulation study. In this section I give special attention to bootstrap procedures that are adequate to deal with structural breaks and unit roots in time series. Also in this section, I make a modest contribution by modifying standard bootstrap procedures to deal with a unique issue that is prevalent when dealing with unit roots and structural breaks. The results from the simulations are reported in section 3.6. This section explains the results on both inference and cointegration testing. Section 4.4 provides a practical guide on how to deal with unit roots and structural breaks. I provide a stepwise discussion on how to approach the testing of unit roots and ensure accurate inference in cointegrated regressions. Section 3.7 concludes this chapter.

In Chapter 4, I demonstrate the insights of the simulation study in chapter 2, by applying alternative techniques to data from a European competition case, providing guidance for practitioners on the strengths and weaknesses of alternative approaches. Section 4.2 provides an overview of the case and section 4.3 reports the results. In section 4.4, practical guidance is provided on how to deal with unit root testing and possible modelling solutions are discussed. This is supported by results that is presented in section 4.4.1. Section 4.5 summarizes the chapter.

Chapter 5 is the final chapter of this dissertation. In it, I provide a summary of the research questions, results and practical implications that are outlined in Chapters 2, 3 and 4. This section concludes with a discussion of the limitations of this thesis and remaining unanswered research questions. I provide my intuition on how to approach these questions in the hope of guiding future research.

Chapter 2

Playing Cupid for cartel dating: The robustness of cartel dating methods

2.1 Introduction

Cartels that engage in price fixing have been an ongoing concern for antitrust enforcement. Despite the illicitness of collusive behaviour, antitrust authorities continue to uncover and prosecute firms that participate in collusion. Once uncovered, the firms involved in collusion are expected to pay penalties based on the value of damages caused by their infringement. Additionally, customers that paid overcharged prices can bring forth civil litigation against cartel members in private actions to obtain reparation.

Private actions, in combination with public enforcement, have been an important component in deterring cartel behaviour. Private actions is also the mechanism through which injured parties can be compensated for damages. When a case is being evaluated, the compensation awarded to the plaintiff will be based on the alleged overcharge. Given that cartels vary in scope, size and their effect on prices the overcharge is estimated on a case-by-case basis.

In recent years, econometric models have become the standard approach for estimating overcharges (European Commission, 2005, 2008; McCrary and Rubinfeld, 2014). Typically, these models seek to estimate a counter-factual scenario with which the factual outcome can be compared. The two most commonly used econometric approaches involve using benchmarks and yardsticks. The benchmark approach evaluates prices in the market in question by comparing prices before and/or after the infringement period with prices during the infringement period. The benchmark approach therefore seeks to establish a counter-factual price that would have prevailed in the absence of collusion. The yardstick approach compares prices in the infringing market with a comparable market where no violation occurred. This chapter is focused mainly

on the benchmark approach.

There are two methods used in typical benchmarks, namely the dummy variable approach and the forecasting approach. Both methods rely on time series data. In the forecasting approach, a model is estimated using prices outside of the infringement period. Using the estimated model, forecasts are then constructed to create a counter-factual "competitive" price. The overcharge is then taken as the difference between the forecast price and the actual collusive price. In the dummy variable approach, data spanning over the entire sample is used and a dummy variable is included to capture the shift in price during the cartel period.

In both benchmark approaches, knowledge is required regarding the start and end dates of the cartel. It is important to know this with some precision in order to distinguish periods of collusion from periods of competition. In practice, antitrust authorities typically rely on evidence given through leniency programmes, or documentary evidence to determine both the start and duration of the infringement period. This evidence often relates to communication between the colluding firms. When determining the collusive period in this manner, the implicit assumption is that the anti-competitive effects began and ended at the exact date on which communication between cartel members started and ended. This need not necessarily be the case since the cartel can take time to reach full effectiveness and, after their disbandment, the anti-competitive effects can take time to dissipate. Examples include markets that are characterised by medium- and long-term contracts and cases where the members' familiarity with one another causes the anti-competitive coordination to continue tacitly, without direct contact with one another. Furthermore, cartels are not always stable over time, with some members defecting. Therefore, collusive periods can sometimes be interspersed by periods of competitive price wars.

The applied literature has recognized the importance of statistically determining the start and end dates of the effective cartel periods. This process is referred to as cartel dating. Cartel dating involves uncovering unknown break dates in price data. Various alternative techniques, including structural break tests and regime-switching models, have been proposed to determine the effective cartel dates. Once the effective cartel dates are established the next step is to use this information to estimate the overcharge. When using the dummy variable approach, the construction of the dummy variable is dependent on the results of the cartel dating procedure. This dependence is not formally discussed in the cartel dating or overcharge literature.

This chapter aims to contribute to the current literature along three dimensions:

> First, any form of structural break test carries the risk of type I and II errors. It is therefore important to understand how these errors can translate into overcharge estimation when structural break tests

are used to inform the construction of a dummy variable. There is substantive econometric literature that considers the probabilities of type I and II errors for various structural break tests[1]. Interestingly, the issue of how these errors map onto parameter estimates – when a dummy variable was constructed based on the structural break test results – has not yet been investigated. The first aim of this chapter is to bridge this gap.

The second dimension is related to the similarities between structural break tests and regime-switching models. These are the two approaches that have been proposed to deal with the issue of cartel dating. Both methods are used to deal with economic relationships that are not constant over time. Surprisingly, the comparative performance of these two approaches has received little attention. The second aim of this chapter is to provide further insights on how structural break tests and regime-switching models compare when used in overcharge estimation.

The third dimension is related to the fact that various methods that deal with structural breaks are typically studied in isolation. That is, the literature evaluates each structural break test or regime-switching model on its own without comparing the results with various alternatives. There is, therefore, little evidence on how various methods perform relative to one another. The third aim is to provide a typical horse-race to determine which method will provide the most accurate overcharge estimation.

The structural breaks in prices caused by collusion take various forms. Structural change may not involve only single hard and sudden shifts at the start and end dates during which the cartel was active. This chapter, therefore, consider the comparative performance of various techniques for different forms of structural break. This allows us to identify the shortcomings of various approaches and the conditions under which they yield similar results.

The rest of this chapter is organized as follows. Section 2.2 discusses the importance of cartel dating in overcharge estimation and the related literature. Section 2.4 provides a brief overview of the various DGPs to which I apply all of the structural break test and regime-switching models. This is followed by the Monte Carlo (MC) design and a technical discussion of each of the structural break tests and regime-switching models. In section 2.5 I present the results of the MC experiments and comparisons between the various approaches. Section 2.6 describes some challenges associated with Markov-switching (MS) models in MC simulation studies with respect to optimization procedures, model specification and software considerations. In section 4.2 I apply the various dating

[1]For comprehensive overviews on this literature see Perron *et al.* (2006) and Casini and Perron (2019)

approaches to a practical collusion case where overcharges have been estimated. Section 2.7 concludes.

2.2 The importance of cartel dating in overcharge estimation

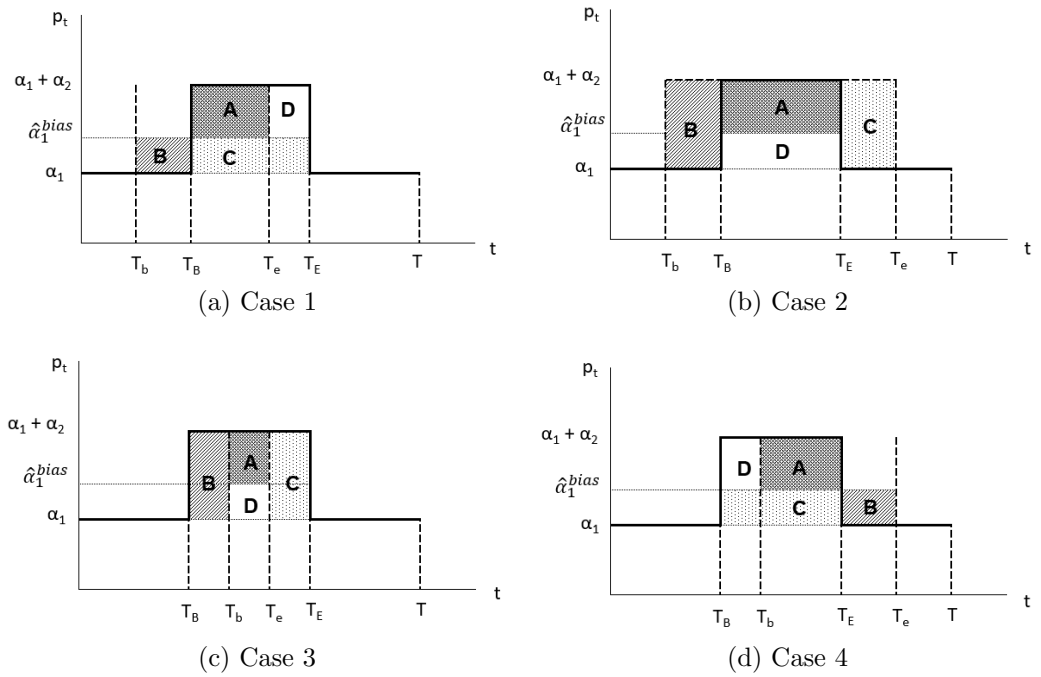
The formal cartel dates determined by the courts and the effective cartel dates, as discussed above, may not coincide. Boswijk *et al.* (2019) investigated the effects of incorrectly specifying the begin and end dates of the collusive effects. The authors show that misdating leads to an underestimation of overcharge. Drawing on Boswijk *et al.* (2019), there are four ways in which a cartel period can be misdated from the actual effective period. Suppose that T_b and T_e are the formal begin and end dates of the cartel, and T_B and T_E are the effective begin and end dates of the cartel. The four possible cases of misdating are:

$$\begin{cases} \text{Case 1: } T_b < T_B, T_e < T_E \\ \text{Case 2: } T_b < T_B, T_e > T_E \\ \text{Case 3: } T_b > T_B, T_e < T_E \\ \text{Case 4: } T_b > T_B, T_e > T_E \end{cases} \tag{2.1}$$

Figure 2.1 graphically illustrates each of the four cases, where α_1 is the competitive price, $\alpha_1 + \alpha_2$ is the collusive price.

In the first case, the cartel is dated to begin and end too early. There is thus a part of the competitive period that is included in the cartel period and a part of the cartel period that is included in the competitive period. This will cause the counter-factual price to be overestimated as $\hat{\alpha_1}^{bias}$ for both the forecasting and dummy variable approaches. For the dummy variable approach the incorrect inclusion of collusive prices in the competitive period will cause overestimation of the counter-factual price. It should be recalled that overcharge is the difference between the estimated counter-factual price and the actual price that was observed. Total overcharge is the overcharge multiplied by the quantity bought from the cartel. There are three reasons why the total overcharge will be underestimated. First, overestimation of the counter-factual price means that area C is omitted. Second, the total overcharge from period T_b to T_B will be negative due to the overestimated counter-factual price, and area B will be subtracted from the damage estimate area A. Third, area D is not included in the total overcharge since T_e lies before T_E.

In Case 2, the cartel is dated to begin too early and end too late. The forecasting approach would use data only from T_0 to T_e and T_e to T. Therefore, no cartel period is erroneously dated as competitive. The counter-factual price will therefore not be biased in this case. However, when using the dummy variable approach the periods T_b to T_B and T_E to T_e are erroneously included



(a) Case 1 (b) Case 2

(c) Case 3 (d) Case 4

Figure 2.1: Cases of misdating

as being collusive when they are, in fact, competitive. This will cause the dummy variable parameter to have downward bias, and hence area D will be missed. When calculating total overcharge using the dummy variable method, the total damage will incorrectly include areas B and C. The direction of bias in the case of total overcharge will therefore depend on the distance between the misdated periods.

In Case 3, the periods T_B to T_b and T_e to T_E are incorrectly labelled as competitive. When using the forecasting approach, the counter-factual price will be overestimated, and area D will be missed. Additionally, when calculating total damage, areas B and C are not included, which means further underestimation of total damage. When using the dummy variable approach the dummy would incorrectly include the periods T_B to T_b and T_e to T_E as competitive. This will cause the dummy variable parameter to be biased downward. Similar to the forecasting approach, when calculating total damage, areas B and C are not include,d and area D is omitted due to the incorrectly estimated counter-factual price.

In Case 4, the cartel is dated to begin and end too late. This is the mirror image of Case 1, where the counter-factual is overestimated and the total damage is underestimated.

In all four cases, the conclusion is that misdating would generally lead to an underestimation of the overcharge and total damages. The implication is that the effective cartel dates need to be used rather than the dates based on qualitative impressions. To determine the effective dates, rigorous econometric

methods are preferable – since they present a unified approach, i.e. the cartel dates are determined using a similar model and framework – to what is used to estimate the overcharge.

2.3 Related literature

The econometric literature features a variety of methods for the identification of structural change. As explained extensively in section 2.4, this literature in includes inter alia the following methods Bai-Perron (BP) break test (Bai and Perron, 1998), cumulative sum (CUSUM) (Ploberger and Krämer, 1992), moving sums (MOSUM) (Chu *et al.*, 1995), and sequential F-tests (Andrews and Ploberger, 1994). In addition, the more recent literature include the test of Qu (2008), Oka and Qu (2011), Perron and Yamamoto (2013), Ciuperca (2014) and Oka and Perron (2018). Qu (2008) and Oka and Qu (2011) consider the identification of multiple structural breaks at unknown dates from one or multiple conditional quantile functions. Oka and Qu (2011) also provides a procedure to determine the number of breaks. Perron and Yamamoto (2013) consider the issue of testing for structural change using a band-spectral analysis. The test allows changes over time to vary within selected frequency bands, permitting the coefficients to be different across the frequency bands. Ciuperca (2014) provides an application using an adaptive Lasso model to determine the timing and number of structural breaks. Oka and Perron (2018) devises a test for common breaks across or within equations in a multivariate system.

I do not consider all available structural break tests in the literature. My choice of a non exhaustive list of methods is informed by legal practice and the scope of the cartel damage literature. While there is likely methods that may outperform the selection, the purpose of this assessment is to compare areas of strengths of the different methods as part of a broader tool-kit. Furthermore, the Monte Carlo simulations differ from those in Antoshin *et al.* (2008) and Groen *et al.* (2013) where the authors consider the ability of the CUSUM and BP test to correctly identify the break dates. The simulation in this dissertation can be viewed as an extension of this literature where I consider how the results from these tests map onto parameter bias.

The applied literature has grappled with the issue of cartel dating, proposing various solutions for different cartel cases. In an earlier attempt Hüschelrath *et al.* (2016) used a sequential dummy variable in combination with t-statistics to find the effective cartel break dates. In this case, the German Federal Cartel Office (*Bundeskartellamt*) found that the cartel was active in the 1990s to the early 2002. However, sequentially testing the dates at which their dummy variable was significant, the author's found that the cartel's effects subsided in November 2001. It should be noted that in this case, data for before the cartel's formation in 1990 was not available, and hence the start

date might also have differed using the same methodology.

Another empirical application of structural break tests to deal with misdating is given in Boswijk *et al.* (2019). The authors used a Bai and Perron (1998) structural break test to determine the start and end periods, and subsequently encoded a dummy variable to represent the structural change in the overcharge estimation. The authors applied the test to the European Sodium Chlorate cartel which, according to the European Commission, operated from September 1994 to February 2000. When applying the Bai-Perron structural break test, the authors found that the effective cartel dates were rather January 1995 to February 2002. This specific example is related to Case 1 in Figure 2.1 (a). Indeed, the authors illustrate that when using the Commission's official dates, the overcharge is less than half of the estimate obtained when using the structural break test determined dates.

A third example of a case where the anti-trust authorities' established dates did not coincide with the cartel effective dates is given in Boshoff and van Jaarsveld (2018). In this paper, the authors evaluated the South African Cement cartel which operated legally from 1986 to 1995 and then illegally from 1998 until 2009. This case presents a unique scenario in which collusion was recurrent. To deal with this phenomenon – as well as the possibility of misdating – the authors employed a Markov-switching (MS) model. While the starting dates of the cartel were consistent with those suggested by the model, the results suggests that prices were significantly lower from 2006 to 2008. Additionally, the cartel effect continued after the official end date of 2009, until 2011. The authors show that when using the official cartel dates, compared with the model-determined dates, the overcharge estimate decreases from 19.89% to 2.12%.

Crede (2019) presents a variety of structural break tests to test for cartel behaviour and determine the start and end dates of a cartel in the pasta industries of Italy and Spain. The formal cartel dates were deemed to be October 2006 to March 2008 for Italy, and July to October 200 for Spain. To evaluate the start and end dates, Crede (2019) employed a variety of different structural break tests, namely, a sequential F test, the OLS-based cumulative sum (CUSUM), moving sum (MOSUM) and the Bai-Perron test. For the cartel, the author again finds that the empirically tested cartel dates differ from the official dates. For both markets the structural-break-determined dates were around June 2007 with an end date of June 2008.

Each of the above-mentioned studies present separate challenges in the applied work that seeks to estimate overcharge. In each case, different methodologies are used to detect the cartel periods. Despite the differences in approach, the main goal remains to determine the timing of structural change in the mean of the dependent price variable. I am therefore not interested in a variety of structural shifts. However, even for this focused set, there is no single encompassing data generating process (DGP) that will hold in every applied case, and the relative performance of various structural change methods – that

exhibits collusive behaviour – is yet to be investigated. In this chapter, I provide insight into this problem. I assess which of the cartel dating methods proposed in Boswijk *et al.* (2019),Boshoff and van Jaarsveld (2018) and Crede (2019) are robust to variations in the DGP and establish the conditions under which they might lead to biased overcharge estimates.

To investigate the performance of various methods of cartel dating, I constructed four different DGPs of cartel prices, each related to documented empirical and theoretical evidence. It is important to stress that the difference in DGPs is related only to the nature of the shift in prices and not to the specific form of misdating. That is, the four DGPs are not related to the various cases of misdating represented in figure 2.1. Hence, any of the four cases of misdating can apply to the various DGPs when relying on qualitative evidence. In this chapter, I focus rather on which methods of cartel dating will give robust results and the conditions under which they give biased estimates.

The first case I consider is the standard cartel case where the cartel raises the average price immediately at a specific date and again lowers the average price immediately at a later date, once the cartel is uncovered. This constitutes the standard assumption that is made when the overcharge is estimated (see European Commission, 2013; Davis and Garcés, 2009; McCrary and Rubinfeld, 2014).

The second case is where the shifts in the mean are driven by an underlying probability process. The work of Lee and Porter (1984) forms the theoretical basis for the formulation of this DGP. The authors used a regime-switching model to evaluate cartel stability. They discuss the issues of using a switching model with a priori information on sample separation (least squares with a dummy) vis-à-vis a model that has exogenous switching (MS). The authors used the Joint Executive Committee (JEC) railway cartel as an example, where the data contained price wars and collusive periods. The paper proves econometrically and illustrates empirically that erroneously assuming that the collusion regimes are deterministic (probability of collusion = 1), while they are not, will lead to a misclassification problem and cause a biased estimate. The theoretical literature also supports the notion that collusion – specifically in the presence of uncertainty – is not always perfect and that the cartel pricing strategy will not always involve pricing at the maximum joint-profit-maximizing level (Porter, 1983; Green and Porter, 1984). In the simulated DGP, by having the mean shifts driven by a probability process, the timing and duration of collusion is uncertain and imperfect (probability of collusion $\neq 1$ for any period).

The third case represents the situation where collusion is recurrent. Recurrent collusion is possible in at least two settings. First, in markets with a legal history of collusion illegal cartels often reappear (Boshoff and Van Jaarsveld, 2019). Second, canonical models of collusion treat collusive outcomes as being state dependent which is usually related to demand (Rotemberg and Saloner, 1986; Haltiwanger and Harrington Jr, 1991). As a result of demand shocks,

collusive periods are often interspersed with price wars. Both factors can, therefore, cause the data under consideration to contain multiple periods of competition and collusion.

The fourth case I evaluated is when a gradual transition exists between collusive and competitive phases. In this case, the cartel is deemed to gradually increase their prices to the optimal profit-maximizing level. The European Commission (2013) recognizes that in practice, this is often the case due to practical reasons or the cartel simply attempting to avoid suspicion. In the empirical application, Hüschelrath *et al.* (2016) state that based on qualitative evidence, the German cement cartel might have behaved in this manner. An additional reason for the phasing out of the effective period, after the cartel's disbandment, is lingering effects. This can occur for a variety of reasons related to tacit collusion following explicit collusion, uninformed buyers, reduced capacity, existence of long-term contracts, and strategic behaviour to reduce penalties (Harrington Jr, 2004).

The goal of this chapter is to test which structural break methods proposed in the cartel overcharge literature is robust to the various forms of collusive price behaviour. Specifically, the chapter is focused on which methods can correctly determine the start and end points of structural change in the mean. The applied cartel overcharge literature proposed five different approaches: the sequential testing procedure, Bai-Perron structural break test, CUSUM, MOSUM and an MS model. I, therefore, applied all five methods and additionally applied a time-varying parameter model (TVP) to each of the DGP's discussed previously. I relate the structural break tests accuracy directly to the overcharge estimation. In other words, I assessed each method's ability to correctly estimate overcharge when the break dates suggested by the test were used to determine the overcharge.

I make the following contributions to the literature. First, the empirical performance of structural break tests is typically evaluated separately (for an overview see Perron *et al.*, 2006). In this chapter I compare the relative performance of various methods in determining structural change where the change was driven by different dynamics. Second, I extend the typical framework and evaluate how the various structural break tests performed when the results were used to construct dummy variables that controlled for the corresponding breaks. Third, to the best of my knowledge, there is no literature that compares the performance of structural break tests with that of regime-switching models. Given that structural break test are often used to motivate various controls that capture the effects it is important to understand the similarities and trade-offs between the two approaches. Additionally, the results provide guidance for practitioners in overcharge cases.

While the main emphasis was on overcharge estimation, the results in this chapter are generally relevant. As explained with Figure 2.1, misdating periods of structural change will lead to a biased estimate. This is relevant to the applied econometric literature, where dummy variables are frequently used to

control for structural change (Wooldridge, 2016, 344). The results show that care needs to be taken with regard to the nature of the structural change and the method applied to determine the periods thereof. Furthermore, I illustrate the interplay between structural break tests and a model that can accommodate structural change. The results show that despite the similarities of the two approaches, the results will not always be similar.

2.4 Methodology

To incorporate changes that are at unknown dates into an econometric model, there are two approaches that can be followed. One approach is to use structural break tests to determine the start and end dates of the change in the mean, and subsequently to use these dates as a guideline to encode a dummy variable that represents the change in mean. Thereafter, the dummy can be included as an exogenous regressor in the model estimation. Alternatively, structural change can be modelled directly by using a regime-switching model that accounts for different means between regimes.

This chapter is focused on evaluating various approaches to detect and subsequently model changes in the mean of a DGP. Understanding the properties of different approaches is vital to cartel damage estimations. To this end, I considered four different types of shifts in the mean, as discussed in section 2.2. Each of these cases represents a different form of mean shift that has been reported in cartel cases or suggested in the cartel literature.

The choice of break tests and regime-switching models are motivated by what is standard practice in the applied literature as well as by what has been used in the cartel damage literature. I considered the following structural break tests: the Bai-Perron break test (BP), cumulative sum (CUSUM), moving sums (MOSUM), and F-tests. For regime-switching models I consider MS and TVP models. More specifically, I used a Markov-switching dynamic regression model (MSDR). The MSDR specification is chosen since it allows for quick adjustments between regimes and the regime-switching model parameters are independent of the previous regimes. The MSDR specification is discussed in more detail in section 2.4.2.5.

To investigate the performance of each break test and regime-switching model, I performed Monte Carlo simulations for each of the four cases of shifts in the mean. That is, I simulated thousands of values from a given DGP and, for each iteration, performed the break test, estimated the model and then recorded the result. I then performed a variety of summary statistics on the results to assess the bias from each method. The following section explains the Monte Carlo simulations in greater detail and explains the break tests and model estimations.

In section 2.4.1 I explain the Monte Carlo simulations and four different DGPs that were evaluated. Section 2.4.2 provides a brief technical discussion

of each structural break test and model estimation. In section 2.4.3 I explain how I calculated the estimation bias and compare the Monte Carlo results.

2.4.1 Monte Carlo simulations

In econometrics, Monte Carlo (MC) simulations are frequently used to explore the properties of various estimators, tests statistics, and the derivation of non-standard distributions of test statistics (Hamilton, 1994, 217; Enders, 2008, 202). For regression analysis, estimators and test statistics are constructed from a random sample of data. The random behaviour of estimators and test statistics are described by their sampling distribution. In classical frequentist approaches, the sampling distribution is of utmost importance. Knowledge of the sampling distribution forms the basis for measuring the accuracy of the tests and estimators. Furthermore, when the sampling distribution is unknown, the appropriate critical regions will also be unknown.

When estimating regressions it is important to avoid estimators that are inconsistent. One of the implications of consistent estimators is that the estimator will be asymptotically unbiased – that is unbiased when the sample is infinitely large. However, when only a limited sample is available there is no guarantee that the estimators will be unbiased. Unfortunately, there is no criterion for an optimal sample size that will guarantee that the asymptotic results will hold.

To evaluate the performance of test statistics and estimators there are two important considerations. First, I need to know the sampling distribution. Second, to obtain the sampling distribution, I require an infinite number of samples from the population. This is where Monte Carlo simulations become useful. Monte Carlo simulations are based on the *law of large numbers* and *central limit theorem*. The basic premise is to replicate the sampling distribution. In principle, this is done by drawing a random sample from the population, constructing the test statistic of interest and repeating the process a large number of times. The values of the test statistic in each of the repeated simulations together with the frequency of occurrence is what makes up the approximate sampling distribution.

Given the above background, practical data is not suitable for evaluating the consistency of test statistics and estimators for the following reasons. First, to determine bias in small samples, I require knowledge of the sample distribution. When using actual data that do not constitute the full population, this cannot be stated with absolute certainty. Second, to determine if a test statistic or parameter is biased, I need to know the exact functional form and specification of the DGP. Without this precise knowledge, it will be impossible to determine if the bias or test error resulted from the misspecification of the DGP. Therefore, to test the hypothesis – on the performance of various cartel dating methods – I make use of MC simulations.

The rest of this subsection explains the different DGPs for which I performed the MC simulations. I also motivate my choice of parameter values and provide graphical illustrations of the typical DGP in each case.

Suppose that price over time develops according to the following DGP.

$$p_t = \alpha_1 + \alpha_2 D_t + \beta \mathbf{x_t} + \gamma p_{t-1} + \epsilon_t, \tag{2.2}$$

where p_t is the product unit price in period t, D_t is the cartel dummy variable, $\mathbf{x_t}$ is a set of explanatory variables. Referring back to equation 1.1 in chapter 1 this x_t may consist of vectors d_t and c_t as described in equation 1.1. Limiting the terminology to a single vector x_t, instead of two separate cost and demand vectors, enhances the exposition of the simulation results as the results are not influenced by the number of exogenous variables. I recognize that empirical specifications – including the application in chapter 4 – may allow for a higher lag order. However, in initial testing further lags did not significantly alter the results. I assume that $0 < \gamma < 1$, i.e. the dynamic relation is stable.

The sample period is $\mathbb{T} = \{1, ..., T\}$, can be divided into periods with a cartel effect, labelled $\mathbb{T}_C$, and a period without a cartel effect, labelled $\mathbb{T}_N$, hence $\mathbb{T} = \mathbb{T}_C \cup \mathbb{T}_N$. The cartel effect is captured by the dummy variable D_t, which is equal to 1 in the effective cartel periods ($t \in \mathbb{T}_C$) and 0 otherwise ($t \in \mathbb{T}_N$). For simplicity, first consider one single and continuous period of cartel effects with a known begin date T_B and end date T_E, with $1 < T_B < T_E < T$, so that $\mathbb{T}_C = \{T_B + 1, ..., T_E\}$, $\mathbb{T}_N = \{1, ..., T_B, T_E, ..., T\}$ and

$$\begin{aligned} D_t &= 0, \ t \leq T_B \\ D_t &= 1, \ T_B < t \leq T_E \\ D_t &= 0, \ t > T_E \end{aligned}$$

The coefficient α_2 measures the immediate or short-run response of the price to the collusion as a price-level shift. I analysed whether population regression coefficients were accurately estimated, based on a sample of time series observations.

For the Monte Carlo study, the data were generated according to DGP 2.2 with $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$. The cartel dummy was set at $T_B = \frac{1}{4}T$ and $T_E = \frac{3}{4}T$, with $T = 100$ observations. The explanatory variable x_t develops according to an AR(1) model:

$$x_t = \rho x_{t-1} + v_t, \tag{2.3}$$

where $v_t \sim NI(0, \sigma_v^2)$, i.e. the cost factor was assumed to be strictly exogenous. I set $\alpha_1 = 100(1 - \gamma)$, so that the mean price level in the simulations was equal to 100. I set $\alpha_2 = 10$, so that the long run impact of the mean shift in price is 20. I normalized with respect to the variance of the disturbance term σ_ϵ^2. I furthermore chose $\gamma = 0.5$ and $\rho = 0.5$. These values correspond to an intermediate degree of serial correlation in the time series p_t and x_t. To facilitate the comparison of simulation results across experiments, I set β such

that the long-run effect of x_t on p_t was unity, i.e. I specified $\beta = 1 - \gamma$. I choose $\sigma_\epsilon^2 = \sigma_v^2 = 1$.

I chose these parameter values for the following reasons. The increase of a 10% mean is at the lower end of estimates of price effects of collusion. The international literature has found the mean of price overcharges to be 23%, see Connor (2014). In a South African context, estimates of cartel overcharge in the scholarly literature ranges between 16% and 56%. For example, Boshoff (2015) find increases of 25% in bitumen prices, Khumalo *et al.* (2014) finds increases of 16.5% and 57% in concrete pipes, Mncube (2014) find increases between 7% and 42% for wheat flour and Boshoff and Van Jaarsveld (2019) finds 20% for cement prices. For the autoregressive parameter, 0.5 is chosen as a midpoint between nonstationary and stationary. The issue of nonstationarity is explained further in chapter 3. Robustness checks are done for alternative choices in appendix A.

The data was generated recursively. I set the initial value of p_1 as equal to the unconditional mean of the non-cartel period. Due to the auto-regressive dynamics, subsequent realizations of p_t would then also depend on its previous value. To avoid dependence on the first value of p_1 in all the simulated iterations a burn-in period of 100 observations was included. In other words, I simulated $100 + T$ observations and then omitted the first 100. This ensured that the start of each iteration did not contain similar dynamics, which might have biased the results.

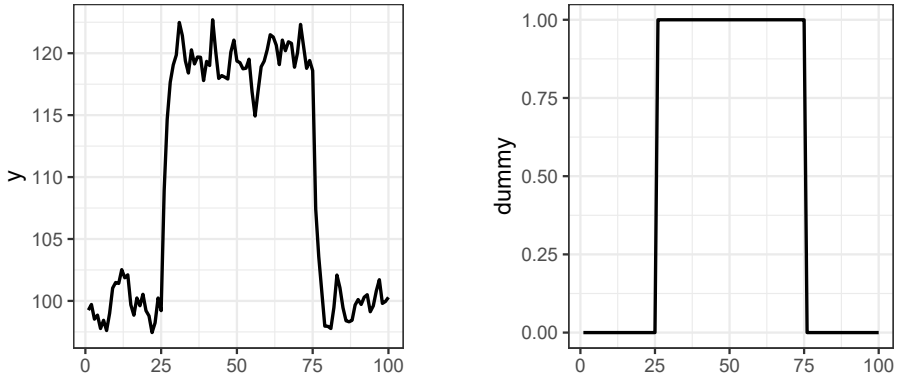
As previously stated, I considered four different types of shift in the DGP. Each case was motivated by the empirical and theoretical literature and encompassed the most likely behaviour that cartel data should exhibit.

Case 1 *The first case considers a single deterministic shift in the mean of the DGP.* The price is driven by equation 2.2, with a dummy variable that evolves according to the following rule-set:

$$\begin{aligned} D_t &= 0, \ t \leq T_B \\ D_t &= 1, \ T_B < t \leq T_E \\ D_t &= 0, \ t > T_E \end{aligned}$$

To provide a clear depiction of the structure of this process, Figure 2.2 plots the price depicted as y and the dummy variable that represents the shift in mean.

Case 2 *The second case represents shifts in the mean of the process that is driven by a Markov process.* Conceptually, the way in which the DGP of 2.2 evolves is not dependent on an underlying probability process that is driven by some variable. The probability process in 2.2 can, therefore, be thought of as degenerate in some sense since th mean shifts occur at fixed points in time. The MS model assumes that there is a probability law governing the



(a) Simulated values of y (b) Dummy variable encoding

Figure 2.2: Example of a DGP simulation with deterministic changes

shift. More specifically, the probability law is assumed to follow a first-order Markov chain. By simulating the DGP according to 2.2 this property has not been incorporated.

Therefore, I, extended the simulations to evolve according to a first-order Markov chain. Thus let S_t denote a discrete-value state variable that denotes the regime in operation at time t. The probability law that governs the value of S_t is assumed to follow a two-regime first-order Markov chain, with the following transition matrix:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi(S_t = 1|S_{t-1} = 1) & \xi(S_t = 2|S_{t-1} = 1) \\ \xi(S_t = 1|S_{t-1} = 2) & \xi(S_t = 2|S_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix}, \tag{2.4}$$

where $\xi(S_t = j|S_{t-1} = i) = \xi_{ij}$ denotes the probability of switching from regime i at time $t-1$ to regime j at time t. For the Monte Carlo simulations, I used the following fixed transition probability matrix:

$$\boldsymbol{\xi} = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix}, \tag{2.5}$$

The state S_t was recursively simulated to create a Markov chain, with initial state $S_1 = 1$. State 2 was assumed to be the cartel effective state, and with $\xi_{11} > \xi_{22}$ there was a higher probability for the DGP to remain in $S_t = 1$. Therefore, there would be fewer cartel effective regimes than normal regimes. Since $\xi_{22} = 0.8$, the DGP would remain in $S_t = 2$ for an extended duration. The fixed transition probabilities were therefore persistent, and S_t is an ergodic[2] Markov chain.

[2]This means that if the process enters a certain state, it can still transition to a different state. Non-ergodic, for example, would be if $\xi_{22} = 1$ and $\xi_{12} \neq 0$,.

There are two main MS specifications, namely Markov-switching autoregressions (MSAR) and Markov-switching dynamic regression (MSDR). In section 2.5.2 I report the results of the MSDR specification. However, in the discussion of the model specification and software usage (section 2.6.1), I simulated data using the MSAR specification.

For the MC simulations of the MSDR model I simulate the DGP as:

$$p_t = \begin{cases} \alpha_1 + \beta x_t + \gamma p_{t-1} + \epsilon_t \text{ for } S_t = 1 \\ \alpha_1 + \alpha_2 + \beta x_t + \gamma p_{t-1} + \epsilon_t \text{ for } S_t = 2 \end{cases}, \tag{2.6}$$

where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$. Equation 2.6 can be re-written as $\alpha_{S_t} + \beta x_t + \gamma p_{t-1} + \epsilon_t$ where $\alpha_{S_t} = \alpha_1$ for $S_t = 1$ and $\alpha_{S_t} = (\alpha_1 + \alpha_2)$ for $S_t = 2$.

The focus of the simulation is on the comparative ability of different methods to estimate α_2. I assume that in evaluating and comparing the different methods that I know the correct functional form. If I were to alter the specification of 2.6, to include lags of x_t for example or additional lags of p_t, that would also alter the functional forms that I feed to the different methods. There is no a priori reason to expect any different conclusions.

For the MC simulations of the MSAR model I simulate the DGP as:

$$p_t = \alpha_{s_t} + \beta x_t + \gamma(p_{t-1} - \alpha_{s_{t-1}} - \beta x_t) + \epsilon_t, \tag{2.7}$$

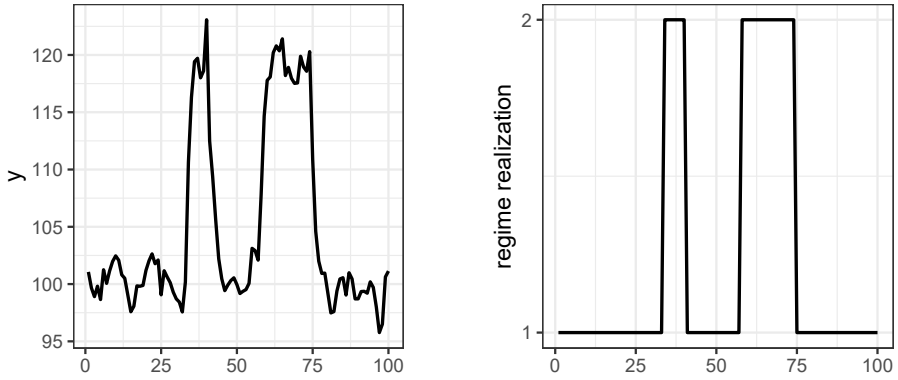
where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$, α_{s_t} represents the regime dependent mean shift, and $\alpha_{s_{t-1}}$ is the intercept of the previous state.

The key difference between the specification in 2.6 and 2.7 is that p_t depends on the current state only in 2.6 while it depends on the current state as well as the state in the previous period in 2.7. I deal extensively with this matter in section 2.6.1 and recognize that the terminology is potentially misleading as both equations are dynamic and autoregressive in nature.

It is important to note that the number of changes in regimes, length of the changes and the timing thereof will not be consistent across simulation iterations. This is due to the fact that changes in regimes are driven by a probability law and I therefore cannot restrict these properties, since doing so would remove the fundamental property of having the changes in mean driven by a probability process.

Providing an example of what these processes might look like, Figure 2.3 illustrates the changes in price represented by y as well as the regime realizations that were driven by a Markov chain.

Case 3 *The third case contains recurrent shifts in the DGP.* In this scenario, I considered two shifts in the mean that are both of equal size and length. This case followed the same DGP as given in equation 2.2 with the following



(a) Simulated values of y (b) Regime realizations

Figure 2.3: Example of a DGP simulation from an MS process

dummy variable:

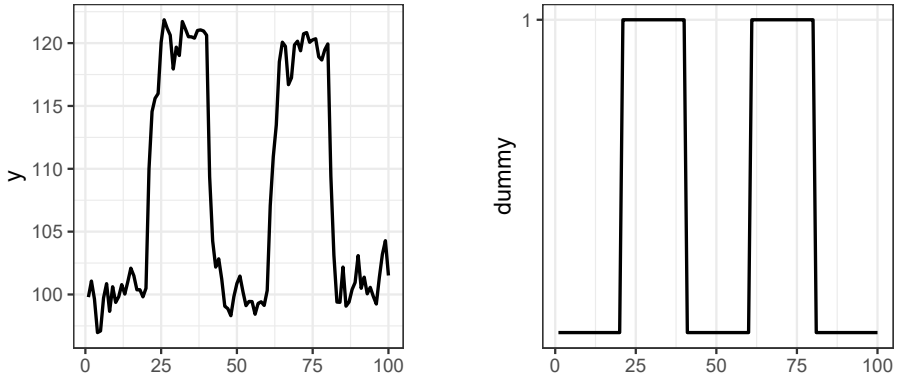
$$\begin{aligned} D_t &= 0,\ t \leq T_B \\ D_t &= 1,\ T_{B_1} < t \leq T_{E_1} \\ D_t &= 0,\ T_{E_1} < t \leq T_{B_2} \\ D_t &= 1,\ T_{B_2} < t \leq T_{E_2} \\ D_t &= 0,\ t > T_{E_2}, \end{aligned}$$

where the begin and end dates of the two cartel periods were defined as T_{B_1},T_{B_2},T_{E_1} and T_{E_2} respectively, with $1 < T_{B_1} < T_{E_1} < T_{B_2} < T_{E_2} < T$. Hence the cartel period ($\mathbb{T}_C$) and non-cartel period ($\mathbb{T}_N$) is $\mathbb{T}_C = \{T_{B_1} + 1, ..., T_{E_1}, T_{B_2} + 1, ..., T_{E_2}\}$ and $\mathbb{T}_N = \{1, ..., T_{B_1}, T_{E_1} + 1, ..., T_{B_2}, T_{E_2}, ..., T\}$. The time periods in this case are set as $T_{B_1} = T_{E_1} = T_{B_2} = T_{E_2} = \frac{1}{5}T$. For a graphical representation of this scenario, see Figure 2.4.

Case 4 *The fourth case involves a transition phase.* In this case, the dummy variable took some time to reach full effect and the effect took time to fade out. The DGP followed is similar to that of equation 2.2 but has the following dummy variable:

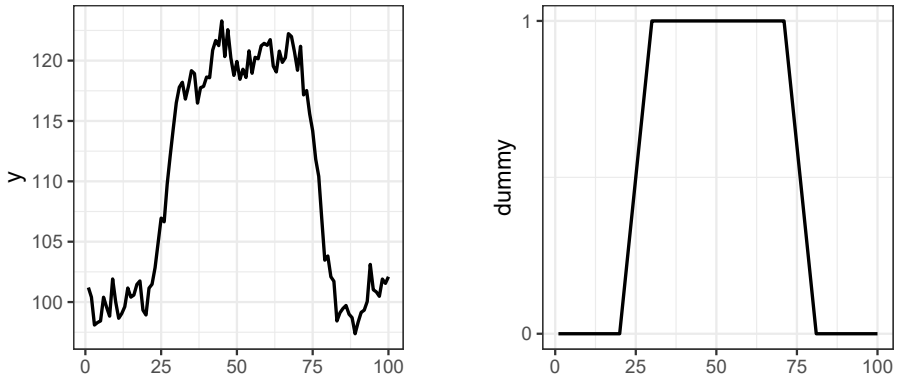
$$\begin{aligned} D_t &= 0,\ t \leq T_B \\ D_t &= \{0, 1, 0.2, 0.3, ..., 1\},\ T_B < t \leq T_{P_B} \\ D_t &= 1,\ T_P < t \leq T_E \\ D_t &= \{1, 0.9, 0.8, ..., 0\},\ T_E < t \leq T_{P_E} \\ D_t &= 0,\ t > T_{P_E} \end{aligned}$$

with $1 < T_B < T_{P_B} < T_E < T_{P_E} < T$. An example of what this case would look like in the simulation is given by Figure 2.5



(a) Simulated values of y (b) Dummy variable encoding

Figure 2.4: Example of a DGP simulation from a recurrent process



(a) Simulated values of y (b) Dummy variable encoding

Figure 2.5: Example of a DGP simulation with transition phases

2.4.2 Model estimations and break tests

I evaluated the simulated breaks using four alternative structural break tests and two regime-switching models. For the break tests, I first tested each simulation to determine the break dates. The break dates were then used as a guide to encode a dummy variable, with the break period represented by 1, and 0's elsewhere. This break-test-informed dummy variable was then used in the following least squares regression:

$$p_t = \alpha_1 + \alpha_2 D_{t,break} + \beta x_1 + \gamma p_{t-1} + \epsilon_t, \tag{2.8}$$

where $D_{t,break}$ represents the dummy variable that was constructed as a result of the break test. This equation specification is identical to the simulation DGP of equation 2.2. Hence, if the break test picked up the correct break dates, the

subsequent model estimation should obtain similar estimates to those specified in the simulation of equation 2.2.

For the two regime switching models, I estimate these models with a change in intercept and recorded how closely the parameter values and changes in intercept between regimes mimiced the behaviour of the simulated DGP. In other words, for each simulation I fitted a regime-switching model with a specification similar to that of equation 2.2, and subsequently recorded the parameter estimates for each run.

The following subsections first discuss the various break tests and then explain the model estimations.

2.4.2.1 Bai-Perron

Bai and Perron (1998) provide the foundations to use the least squares principle to estimate a structural change model. The estimation starts with the following multiple linear models that incorporates m breaks:

$$\begin{aligned} y_t &= \beta x_t + z_t\delta_1 + v_t \ , \ t = 1, ..., T_1 \\ y_t &= \beta x_t + z_t\delta_2 + v_t \ , \ t = T_1 + 1, ..., T_2 \\ &... \\ y_t &= \beta x_t + z_t\delta_1 + v_t \ , \ t = T_{m+1} + 1, ..., T_1 \end{aligned} \tag{2.9}$$

where y_t is the endogenous variable, x_t and z_t are exogenous variables and v_t is the disturbance term. There are $m+1$ regimes which implies m break points. The break points $(T_1, ..., T_m)$ are treated as exogenous and estimated along with the parameters β and δ. Equation 2.9 can be simplified by representing it in matrix notation:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \bar{\boldsymbol{Z}}\boldsymbol{\delta} + \boldsymbol{V} \tag{2.10}$$

where $\boldsymbol{Y} = (y_1, ...y_T)$‘, $X = (x_1, ..., x_T)$‘ and $\boldsymbol{Z}$ partitions diagonally at the m partition $(T_1, ..., T_m)$ that is $\bar{\boldsymbol{Z}} = diag(Z_1, ..., Z_{m+1})$ with $\boldsymbol{Z_i} = (z_{T-i}, ..., Z_{T_i})$‘, $\boldsymbol{\delta} = (\boldsymbol{\delta_1'}, ..., \boldsymbol{\delta_{m+1}'})$ and $V = (v_1, ..., v_T)$.

Using least squares regression, the set of parameters is obtained by minimizing the sum of squared residuals:

$$\left(\boldsymbol{Y} - \boldsymbol{X\beta} - \bar{\boldsymbol{Z}}\boldsymbol{\delta}\right)' \left(\boldsymbol{Y} - \boldsymbol{X\beta} - \bar{\boldsymbol{Z}}\boldsymbol{\delta}\right) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}}^{T_i} \left[y_t - x_t'\beta - z_t'\delta\right]^2 \tag{2.11}$$

If the number of partitions (m) is given the estimates can be denoted as $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$. Let $S_T(T_1, T_2, ..., T_m)$ denote the sum of squared residuals given by the estimates of $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$. The estimated break dates will then satisfy:

$$\left(\hat{T}_1, ..., \hat{T}_m\right) = arg \underset{(T_1,...,T_m)}{min} S_T(T_1, ..., T_m) \tag{2.12}$$

Hence, $\hat{\beta}(\{T_j\})$ and $\hat{\delta}(\{T_j\})$ and each partition are global minimizers of the objective function. The minimization over each of the partitions $(T_1, ..., T_m)$ is such that $T_i - T_{i-1} \geq [\varepsilon]$. ε is defined as a small positive value such that: $\varepsilon : \Lambda_\varepsilon = \{(\lambda_1, ..., \lambda_m); |\lambda_{i+1} - \lambda_i| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq 1 - \varepsilon\}$ where λ_i for $i = 1, ..., m$ represents the break fraction $\lambda_i = \frac{T_i}{T_m}$. This set of arbitrarily small numbers is defined in order to bound the break dates from the boundaries of the sample and to restrict each break date to be asymptotically distinct.

To estimate the period specific coefficients together with the m break points is computationally burdensome. Bai and Perron (2003*a*) show that a dynamic programming algorithm, based on the Bellman principle, reduces the computational burden significantly. Briefly, the procedure can be described as follows: Let the sum of squared residuals with the optimal partition that contains r breaks and n observations be denoted as $SSR(\{T_{m,T}\})$. According to the Bellman principle, with the optimal partitions in hand the following recursive problem is solved:

$$SSR(\{T_{m,T}\}) = \underset{m=h \leq j \leq T-h}{min} [SSR(\{T_{m-1,j}\}) + SSR(j+1, T)] \tag{2.13}$$

where $SSR(j, i)$ is calculated by using the recursive residual $v(i, j)$ such that $SSR(i, j) = SSR(i, j-1) + v(i, j)^2$.

To solve this dynamic problem Bai and Perron (2003*a*) first find the optimal one-break partition and sequentially repeat the process until the optimal $m-1$ partitions are found. The critical values for their test are given in Bai and Perron (2003*b*).

The MC results are obtained by performing the Bai-Perron test (BP) on the estimated regression that has the following form:

$$p_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\gamma} p_{t-1} + \hat{u} \tag{2.14}$$

The break dates from the test are then used to construct a dummy variable where 1s are used to represent the mean shift and 0s are used elsewhere. Subsequently, the BP-informed dummy variable is then used with the following regression specification:

$$p_t = \hat{\alpha_1} + \hat{\alpha_2} D_{T,BP} + \hat{\beta} x_t + \hat{\gamma} p_{t-1} + \hat{u} \tag{2.15}$$

Note that if the BP test correctly identifies the break dates, the resulting estimates should not be different from the parameter values used in the simulation since the specification of equation 2.15 will be exactly the same as that used in 2.2. This process is repeated thousands of times to evaluate the performance of the BP-constructed dummy variable parameter.

2.4.2.2 CUSUM

Given a linear regression model, the CUSUM process is a fitted empirical fluctuation process constructed from the cumulative sums of standardized residuals. Brown *et al.* (1975) were the first to develop a test based on the cumulative sums of recursive residuals. They consider the following CUSUM process:

$$W_n(t) = \frac{1}{\sigma\sqrt{\eta}} \sum_{i=\kappa+1}^{\kappa+|t\eta|} \tilde{u}_i \ (0 \leq t \leq 1), \tag{2.16}$$

where $\tilde{u}_i$ represent the standardized residuals from the fitted regression and $\eta = n - \kappa$ is the number of recursive residuals and $|t\eta|$ is the integer part of $t\eta$. $W_n(t)$ follows a Wiener process or Standard Brownian Motion under the null hypotheses of the limiting process. If the parameter estimates are to remain stable from period to period then $E[W_n(t)] = 0$. However, if the parameter estimates change from period to period, there will be a change in $W_n(t)$ causing it to diverge away from the zero mean line. In other words, if t_0 represents a single change point, then after the point t_0 the recursive residuals will diverge from 0. Therefore, the expectation is that the empirical fluctuation process $W_n(t)$ will be close to 0 up to point t_0.

The basing of a structural change test on the CUSUM residuals was first introduced by Ploberger and Krämer (1992). They define the empirical fluctuation process by:

$$W_n^0(t) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{i=1}^{|tn|} \hat{u}_i \ (0 \leq t \leq 1) \tag{2.17}$$

The limiting process of $W_n^0(t)$ is a Brownian bridge $W_n^0(t) = W(t) - tW(1)$. If there is a single break the path will peak around t_0.

To derive significance bands for the empirical fluctuation process it is important to note that the Brownian motion and Brownian bridge are non-stationary. One approach is to use confidence intervals that represent the standard deviation function which corresponds to the theoretical process. However, linear boundaries can be obtained since a closed-form solution exists for the crossing probability.

For the results in the MC simulations I estimate the following regression on each of the simulated iterations:

$$p_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\gamma} p_{t-1} + \hat{u} \tag{2.18}$$

The empirical fluctuation process is then computed using equation 2.17 and the residuals from the fitted regression. Note that the estimated equation 2.18 is similar to the "true" simulated DGP of equation 2.2 with the only difference being the omitted dummy variable that represents the mean shift. The idea is, therefore, that the empirical fluctuation process will deviate from the null

line at the periods where the mean of the process shifted. After obtaining the CUSUM estimates a dummy variable is created with 1 representing the periods where the process exceeded the significance bands and 0 elsewhere. The dummy variable is then used in a regression with specification exactly the same as that of equation 2.2. To evaluate the process I record the parameter estimates from the regression with the CUSUM-informed dummy variable. This process is repeated thousands of times, where for each iteration, I obtain the CUSUM estimates using equation 2.18, construct the dummy variable, estimate the regression with the dummy variable and same specification as equation 2.2 and record the parameter estimates.

2.4.2.3 MOSUM

In a similar vein to the CUSUM process, another way to analyse structural changes using residuals is through the analysis of moving sums of residuals. Instead of containing the sum of all the sums of residuals up to time t the MOSUM process contains the sum of a fixed number of residuals for a window size determined by the bandwidth parameter h which is moved over the entire sample period. The MOSUM process is therefore defined as:

$$M_n(t|h) = \frac{1}{\sigma\sqrt{n}} \sum_{i=\kappa+[N_n t]+1}^{\kappa+[N_n t]+[nh]} \tilde{u}_i \ (0 \leq t \leq 1-h) \tag{2.19}$$

$$= W_n\left(\frac{[N_n t]+[nh]}{n}\right) - W_n\left(\frac{[N_n t]}{n}\right) \tag{2.20}$$

with $h \in (0,1)$ and $N_n = (n-[nh])/(1-h)$. The OLS-based MOSUM process is, therefore, defined by

$$M_n^0(t|h) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{i=\kappa+[N_n t]+1}^{\kappa+[N_n t]+[nh]} \hat{u}_i \ (0 \leq t \leq 1-h) \tag{2.21}$$

$$= W_n^0\left(\frac{[N_n t]+[nh]}{n}\right) - W_n^0\left(\frac{[N_n t]}{n}\right) \tag{2.22}$$

It follows from equation 2.19 and 2.20 that the MOSUM processes are increments of a Brownian motion and Brownian bridge. This result is discussed in greater detail in Chu *et al.* (1995).

For the MC simulations I follow the same procedure as described in section 2.4.2.2. That is I obtain the break dates using the MOSUM estimate, encode a dummy variable with 1s constituting the break period and 0s elsewhere. Thereafter I estimate a regression with the same specification as equation 2.2. This process is then repeated each time saving the parameter estimates from the regression that contains the MOSUM informed dummy variable.

2.4.2.4 Sequential F test

An alternative to evaluating empirical fluctuation processes is to have a formal statistical test with the null hypotheses of "no structural breaks". Using F-test statistics, Chow (1960) introduced a test where the break dates are known. The test relies on estimating two regressions for the sub-samples where one sample spans the period before the break and another the period after the break. The test statistic is then constructed according to:

$$F = \frac{[SSR_p - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \cdot \frac{[n - 2(k+1)]}{k+1} \tag{2.23}$$

where SSR_1 is the sum of squared residuals for the one sample estimate and SSR_2 the sum of squared residuals for the other sample. SSR_p is the sum of squared residuals for the entire sample obtained by pooling the groups, n is the number of observations and k the number of restrictions.

The obvious drawback of this procedure is that the break date needs to be known. One way to address this drawback is to perform the test by using a rolling window. However, performing this test multiple times over the sample period will increase the probability of committing a type I error for some of the tests. The longer the sample, the higher the probability of committing type I errors. To lower the probability of a type I error, Andrews (1993) and Andrews and Ploberger (1994) suggested various test statistics.

To consider the proposed test statistics suppose that all potential change points are contained in the interval $[\underline{i}, \bar{i}]$. The F-statistics are then computed sequentially in a rolling window for $k < \underline{i} \leq i \leq \bar{i} < n - k$. The test statistics are then given by:

$$supF = \sup_{\underline{i} \leq i \leq \bar{i}} F_i \tag{2.24}$$

$$aveF = \frac{1}{\underline{i} - \bar{i} + 1} \sum_{i=\underline{i}}^{\bar{i}} F_i \tag{2.25}$$

$$expF = log\left(\frac{1}{\underline{i} - \bar{i} + 1} \sum_{i=\underline{i}}^{\bar{i}} e^{0.5F_i})\right) \tag{2.26}$$

For the statistics, the null hypothesis is rejected when the average, maximum, or $expF$ of the F-statistic gets too large. Andrews and Ploberger (1994) derive the asymptotically optimal tests of the $aveF$ and $expF$ test statistics for instances where a nuisance parameter exists under the alternative but not under the null hypothesis. The authors' results are of particular importance in instances where there is a single structural change. I, therefore, consider the $aveF$ and $expF$ test statistics in the MC simulations.

To obtain the MC estimates I follow exactly the same process as described in section 2.4.2.1, 2.4.2.2 and 2.4.2.3. I use the test statistic to inform the

construction of a dummy variable that is used in a LS regression to model the break date.

2.4.2.5 Markov-switching

Markov-switching (MS) models assume that changes in regimes are driven by an underlying stochastic Markov chain. Given that the simulated DGPs are focused on modelling changes in regimes, I specify the following Markov-switching model:

$$p_t = \alpha_{S_t} + \beta x_t + \gamma p_{t-1} + \epsilon_t \tag{2.27}$$

In this specification the intercept is dependent on S_t, a discrete-value state variable, with $S_t = 1$ or $S_t = 2$, depending on the regime that is present at time t. The probability law driving the value of S_t at time t is assumed to follow a two-regime Markov chain with:

$$P(S_t = j|S_{t-1} = i, S_{t-2} = k, ...) = P(S_t = j|S_{t-1} = i) = \xi_{ij} \tag{2.28}$$

where ξ_{ij} is the transition probability of moving from regime i to j. For the application in this dissertation, it is important to note that the transition probabilities in equation 2.27 depends only on the lagged state and not on the other variables in equation 2.27. Therefore, state transitions are exogenous to x_t, p_{t-1}, and ϵ_t. The choice of a fixed transition probability matrix, rather than a time-varying transition matrix, reflects the cartel context. While cartels experience periods of instability collusion tend to be more episodic rather than haphazard. If the transition probabilities in either the simulation or estimation were to be time dependent, the implication is that the price data may enter and exit certain states too often to be comparable to the price data that one typically observes in cartel cases. As discussed earlier, I used the MSDR specification.

Let $\Omega_t = p_t, p_{t-1}, x_t$ denote the collection of all the observed variables up to time t, and let $\theta = (\sigma, \alpha_{S_1}, \alpha_{S_2}, \beta, \gamma, \xi_{11}, \xi_{22})'$ be a vector of population parameters. The observed data has the following conditional log likelihood function:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{T} log f(p_t \mid \Omega_{t-1};\ \boldsymbol{\theta}) \tag{2.29}$$

The starting values of the filter, in other words how to set $P\left(S_{t-1} = j \mid \Omega_{t-1};\ \boldsymbol{\theta}\right)$ to initialize the Hamilton filter, plays an important role in the end result. I discuss the influence of the starting points in greater detail in section 2.6. When S_t is an ergodic Markov chain, the convention in the literature is to set $P\left(S_{t-1} = j \mid \Omega_{t-1};\ \boldsymbol{\theta}\right)$ equal to the unconditional probability $P(S_0 = i)$. The unconditional probabilities are given by

$$P(S_0 = 1) = \frac{1 - \xi_{22}}{2 - \xi_{11} - \xi_{22}} \tag{2.30}$$

$$P(S_0 = 2) = 1 - P(S_0 = 1) = \frac{1 - \xi_{11}}{2 - \xi_{11} - \xi_{22}} \tag{2.31}$$

The conditional likelihood function and the posterior probabilities for S_t are estimated using the same methodology as presented in Hamilton (1989). To obtain the smoothed probability estimates I follow the procedure of Kim (1994) where the joint probability under the Markov process is given by

$$P(S_t = i, S_{t+1} = j|\Omega_T; \boldsymbol{\theta}) = P(S_t = i|S_{t+1} = j, \Omega_T; \boldsymbol{\theta})P(S_{t+1} = j|\Omega_T; \boldsymbol{\theta}) \tag{2.32}$$

$$= \frac{P(S_t = i|S_{t+1} = j, \Omega_t; \boldsymbol{\theta})}{P(S_{t+1} = j|\Omega_t; \boldsymbol{\theta})} P(S_{t+1} = j|\Omega_T; \boldsymbol{\theta}) \tag{2.33}$$

To obtain 2.33 from 2.32, it is important to note that under the correct assumptions, if S_{t+1} is known, the future data in $(\Omega_{t+1}, \ldots, \Omega_T)$ will contain no additional information about S_t. Therefore, by marginalizing the joint probability with respect to S_{t+1}, the smoothed probability in period t is obtained by

$$P(S_t = i|\Omega_T; \boldsymbol{\theta}) = \sum_{j=1}^{2} P(S_t = i, S_{t+1} = j|\Omega_T; \boldsymbol{\theta})$$

$$= \sum_{j=1}^{2} \frac{P(S_t = i|S_t = iS_{t+1} = j, \Omega_t; \boldsymbol{\theta})}{P(S_{t+1} = j|\Omega_t; \boldsymbol{\theta})} P(S_{t+1} = j|\Omega_T; \boldsymbol{\theta}) \tag{2.34}$$

Contrary to the approach required when using break-tests, the MS estimation allows direct modelling of the changes in the mean of the process. However, MS estimations presented several complications – ranging from optimization procedures to specification considerations and MC simulation complexities – that I needed to account for as a result . The complications surrounding the optimization procedures are discussed in section 2.6.2. The choice in specification and software considerations are discussed in section 2.6.1 and results for issues surrounding convergence issues are presented in section 2.6.3.

2.4.2.6 Time-varying parameter

Time-varying parameter (TVP) models assume that parameters change over the sample period. In terms of modelling changes in the mean of the process, this class of estimation is well suited to track the underlying changes. There is, however, a challenge to consider when translating this into practice for cartel damage estimation. Since the parameter estimates are not expected to be

constant over either the non-collusive or collusive period, this model class is not necessarily able to accurately date the collusive period. However, the TVP model can be a useful alternative when court established break dates need to be used to determine overcharge.

Consider the following classical TVP model:

$$y_t = x_t'\beta(z_t) + u_t, t = 1, ..., T \tag{2.35}$$

Here, the non-constant parameters are a function of the smoothing variable z_t such that $\beta(z_t) = \beta_0(z_t), \beta_1(z_t), ..., \beta_d(z_t)$. One can define z_t either as a function of re-scaled time where $z_t = \tau = t/T$ or as an unknown function of a random variable. Thus, one can test the model's coefficient as $\beta(z_t) = f(\tau)$ or $\beta(z_t) = f(z_t)$. In the time series context, the OLS estimation of the estimators are biased. Beck (1983) explains that it is possible that there might be some interdependence between the errors, u_t - OLS may not be efficient. Further, if estimated by OLS, the parameter's confidence intervals would be optimistically small. Should the lagged dependant be included as a regressor, the OLS is likely not consistent.

Given that the model consists of time-varying parameters that vary in a smooth manner[3], then the polynomial of a sufficient degree can be used to approximate the time path of any parameter (Beck, 1983). Hence, I follow Fan and Gijbels (1996) that uses the combination of OLS and the local polynomial kernel estimator[4]. There are two types of time-varying OLS estimators, namely, a local constant kernel, and the local linear method. I follow the local linear method. Assuming that $\beta(z_t)$ is twice differentiable, the estimates are obtained by the following minimisation:

$$arg \min_{\theta_0,\theta_1} \sum_{t=1}^{T} [y_t - x_t'\theta_0 - (z_t - z)x_t'\theta_1]^2 K_b(z_t - z) \tag{2.36}$$

The above minimisation fits a set of weighted local regressions with a bandwidth (b) determined window size. The kernel $K_b(z_t - z) = b^{-1}K[(z_t - z)/b]$ acts as the weights. The general expression of the local linear estimator is:

$$\begin{pmatrix} \hat{\beta}_t \\ \hat{\beta_t^{(1)}} \end{pmatrix} = \begin{bmatrix} S_{T,0}(z_t) & S_{T,1}(z_t) \\ S_{T,1}(z_t) & S_{T,2}(z_t) \end{bmatrix}^{-1} \begin{pmatrix} T_{T,0}(z_t) \\ T_{T,1}(z_t) \end{pmatrix} \tag{2.37}$$

[3]This means that the parameters do not contain large sudden discrete changes from period t to $t + 1$

[4]Several studies, state that a Bayesian estimation approach is the most preferable, see Primiceri (2005) for example. In the context of cartels the Bayesian methods are less appropriate since the choice of priors will be contested. Another estimation strategy is to make use of Maximum likelihood. These performed worse than the TVP model reported in this chapter. However, I do acknowledge that the TVP model is placed at a distinct disadvantage, since the simulated data is not entirely consistent with data where TVP models are typically applied.

with,

$$S_{T,s}(z_t) = \frac{1}{T}\sum_{i-1}^{T} X_i' X_i (z_i - z_t)^s K(\frac{z_i - z_t}{h}) \tag{2.38}$$

$$T_{T,s}(z_t) = \frac{1}{T}\sum_{i-1}^{T} X_i' X_i (z_i - z_t)^s K(\frac{z_i - z_t}{h}) \tag{2.39}$$

These two estimators are considered to be consistent and asymptotically normal. This is true for various types of dependency (i.e x_t, z_t, u_t) that should satisfy the four main assumptions about the type of distribution.

To better predict the dynamics and development of a process over time, I resort to a modified version of eq.2.35 as an auto-regressive process:

$$y_t = \beta_0(z_t)+\rho_1(z_t)y_{t-1}+...+\rho_p(z_t)y_{t-p}+\beta_1(z_t)x_1 t+...+\beta_d(z_t)x_d t+u_t \text{ for } t = 1, ..., T \tag{2.40}$$

I choose the representation of z(t) to be a random variable instead of the rescaled time option (τ). This choice is motivated by the goodness of fit given the results obtained from the simulations. I follow a more simplified version of eq.2.40 as:

$$y_t = \alpha_t + \beta_t x_t + \rho_t y_{t-1} + u_t \tag{2.41}$$

2.4.3 Bias and MC result comparison

The premise of the simulation is to create repeated simulations p_t using the functional form of equation 2.2 with the dummy variable D_t representative of one of the four collusion cases. For each of the simulations I first perform the structural break tests and use the results to create the dummy variable for the regression. Using the structural-break-informed dummy variable, I fit the model with the same form as equation 2.2.

To assess the relative performance of using each structural break test in overcharge estimations I rely on summary statistics, the *relative efficiency*, and calculate the bias of the estimated dummy variable parameter. I also perform Kolmogorov-Smirnov tests (Kolmogorov, 1933) to see if the sampling distributions generated by the MC simulation for each of the approaches is statistically different.

The parameter of interest in this case is α_2, which is the coefficient of the dummy variable where the encoding of the dummy is motivated by the results from the structural break test. I set $\alpha_2 = 10$ in each of the simulations.

Therefore, if $\hat{\alpha_2}$ is an unbiased estimator for α_2 I would expect:

$$\text{mean} = \frac{1}{R}\sum_{r=1}^{R}\hat{\alpha_{2r}} \xrightarrow[\infty]{R} 10, \text{ var} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\alpha_{2r}} - \alpha_2) \xrightarrow[\infty]{R} 0 \text{ and} \tag{2.42}$$

$$\text{bias} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\alpha_{2r}} - \alpha_2) \xrightarrow[\infty]{R} 0 \tag{2.43}$$

where R is the number of repeated simulations in the MC experiment. In this case I do not expect to observe the exact above result for two reasons. First, while R is large (5000 in this case) the simulations cannot be repeated infinitely. Second, the dummy variable's construction is reliant on the results from a structural break test. When relying on test statistics type I errors are unavoidable. That is even when I set the significance level at 1% I would still expect 50 ($0.01 * 5000$) out of the 5000 iterations to be false positives.

To calculate the relative efficiency (RE) I compare the mean square error (MSE) of the model with a dummy variable informed by a structural break test to the MSE of the exact specification from which the data was generated. The reference MSE is therefore the least squares estimation where the dummy variable is specified exactly as in the DGP used for the simulation. The RE is calculated as:

$$MSE = \frac{1}{R}\sum_{r=1}^{R}(\hat{\alpha_{2r}} - \alpha_2)^2 \tag{2.44}$$

$$RE_i = MSE_i/MSE_{true}, \tag{2.45}$$

where i refers to the structural-break-informed dummy variable model and *true* refers to the model with the exact DGP specification as used in the simulation. When $RE > 1$ the implication is that the model is less efficient than the reference model, while $RE < 1$ indicates that the model is more efficient than the reference model.

An important question when comparing results is to know whether the distributions of $\hat{\alpha_2}$ between different models are similar. For example, I would like to know if the $\hat{\alpha_2}$ parameters obtained when the dummy was informed by the Bai-Perron test is similar to the parameters of $\hat{\alpha_2}$ that were obtained when the dummy variable was informed by the CUSUM test. To evaluate this I make use of the two-sided Kolmogorov-Smirnov (K-S) test which is a non-parametric test for differences between distributions. I also apply the K-S test to see if the structural-break-informed dummy variable models differ from the reference model where the dummy has the exact specification as in the DGP that generated the data. A comprehensive explanation of the K-S test statistic is given in section 3.5.3.

I test the regime-switching models in a similar fashion. However, the regime-switching models are not reliant on pre-testing the DGP for structural

breaks. Instead, these models provide a way to simultaneously determine the timing of a structural break and the size of the effect. The expectation would, therefore, be that the regime switching models should perform better than the results from the structural break test informed models due to the elimination of type 1 errors in the pre-testing of the DGP. The results, however, indicate that this is not the case. This discrepancy can be ascribed to the optimization and convergence issues uncovered in the simulations described in section 2.6.

It is important to note that the MC results and bias calculation of the TVP estimates is handled differently than that of the structural break tests and MS model. Since the estimates are not constant over time, I report the mean of the parameters for each of the iterations. Since the intercept will have both a high and low mean estimate I do not report its mean since this will simply be the mean between the two regimes – making the estimation appear biased when it is not. I, therefore, subtract the time-varying estimate from the true mean of the DGP at each point to construct a mean error measure.

The following section presents the results from the MC simulation.

2.5 Results

To obtain the results I perform 5000 MC iterations for four different DGPs. Section 2.4.1 explained each of these four cases in detail. First I simulate data that comes from a deterministic DGP. Here the shift in mean is represented in the standard OLS framework with a dummy variable capturing the shift in the mean. Second, I simulate data that comes from a MS model. In this case, changes in the mean are driven by a Markovian process and are probabilistic in nature. The third and fourth DGP represent the more nuanced cases. The third DGP involves multiple collusion phases while the fourth DGP considers the case where transition from one regime to another takes a longer period of time and is less abrupt.

For each of the four DGPs I computed four results in a Monte Carlo simulation. First, I compiled summary statistics from the MC simulation using a structural-break-informed dummy variable obtained from the various tests and estimated the two regime-switching models. A least squares regression, where the specification was exactly as in 2.46, was used as a benchmark. Second, I calculated the K-S test statistic between each of the simulation results to test if the results generated by the different procedures had statistically significant differences. Third, I calculated the relative efficiency statistic; That is, how efficient each method is compared with the least squares benchmark. The fourth result is the bias calculation. I present the results as a box-plot in each of the respective sections.

Subsection 2.5.1 contains the results when there is a deterministic shift in the DGP. Subsection 2.5.2 reports the results for a probabilistic shift, i.e. a

MS model driven DGP. Section 2.5.3 reports the results of recurrent shifts and 2.5.4 the results for when the DGP has longer transition phases.

2.5.1 Case 1 - Deterministic shift in DGP

In this setting I generated data using the following DGP:

$$p_t = 50 + 10D_t + 0.5x_t + 0.5p_{t-1} + \epsilon_t \tag{2.46}$$

with $\epsilon_t \sim IN(0, 1)$.

To obtain the results I applied the four different structural break tests to the DGP as in equation 2.46 without the dummy variable. The intuition was that the respective structural break tests will find a break in the intercept. I then ran a regression similar to that of equation 2.46 where the dummy variable was encoded according to the structural break tests results. For each of the iterations I record the values of the parameters. I provide four results, namely summary statistics, K-S test statistic, relative efficiency, and box-plots of the bias.

Comparing the results of the structural break tests and regime-switching models is non-trivial. Comparing the OLS result where the break dates are know with the structural-break-informed model is simple, since both specifications have the same parameters I can compare their accuracy. Comparing the structural break models with the regime-switching models is more complicated, since the regime-switching models, with a change in intercept, do not have a comparable dummy variable parameter. To compare the structural-break-informed model results with the regime-switching models I first calculate the mean shift indicated by the model. For the structural-break-informed model, the mean shift is calculated as $\frac{\alpha_1+\alpha_2}{\gamma}$. For the regime-switching models the mean shift is calculated as $\frac{\alpha_2-\alpha_1}{\gamma}$. I then calculate the error as the difference between the estimated mean shift and the *true* mean shift.

Table 2.1 reports the summary statistics for the least squares benchmark model. Tables 2.2 to 2.5 report the results of the estimation when the dummy variable is informed by each of the structural break tests. Tables 2.6 and 2.7 report the results for the MS and TVP models. If a structural break test performs well at identifying the correct break dates, the expectation is that the mean will be close to the true parameter values and the variance will be small.

Table 2.1: LS simulation benchmark when the DPG is deterministic

	α_1	α_2	β	γ
Mean	50.44	10.08	0.50	0.50
Median	50.36	10.08	0.50	0.50
Variance	8.03	0.30	0.01	0.00
Stdev	2.83	0.55	0.09	0.03
Skewness	0.21	0.12	0.02	-0.20
Kurtosis	0.45	0.09	0.60	0.44

Table 2.2: BP LS simulation results when the DPG is deterministic

	α_1	α_2	β	γ
Mean	50.29	10.05	0.50	0.50
Median	50.60	10.09	0.50	0.49
Variance	26.23	0.94	0.01	0.00
Stdev	5.12	0.97	0.09	0.05
Skewness	-5.60	-5.33	-0.27	5.66
Kurtosis	46.51	43.45	0.50	47.23

Table 2.3: CUSUM simulation results when the DGP is deterministic

	α_1	α_2	β	γ
Mean	2.73	1.07	0.35	0.98
Median	2.63	0.85	0.35	0.98
Variance	0.95	0.58	0.05	0.00
Stdev	0.97	0.76	0.22	0.01
Skewness	0.58	1.39	-0.03	-0.65
Kurtosis	1.61	2.36	-0.19	1.81

Table 2.4: MOSUM simulation results when the DGP is deterministic

	α_1	α_2	β	γ
Mean	2.60	-0.06	0.36	0.98
Median	2.52	-0.02	0.36	0.98
Variance	0.91	2.18	0.05	0.00
Stdev	0.96	1.48	0.21	0.01
Skewness	0.48	-0.36	0.01	-0.48
Kurtosis	1.92	2.95	-0.10	1.83

Table 2.5: F-test simulation results when the DGP is deterministic

	α_1	α_2	β	γ
Mean	4.09	0.52	0.37	0.96
Median	4.03	0.53	0.38	0.96
Variance	1.67	0.09	0.05	0.00
Stdev	1.29	0.30	0.22	0.01
Skewness	0.23	0.02	-0.03	-0.19
Kurtosis	0.84	3.40	-0.14	0.72

Table 2.6: MS simulation results when the DPG is deterministic

	α_1	α_2	β	γ	α_2 - α_1
Mean	50.07	60.43	0.50	0.50	10.36
Median	50.03	59.98	0.50	0.50	10.33
Variance	6.69	23.28	0.01	0.00	30.48
Stdev	2.59	4.83	0.09	0.03	5.52
Skewness	0.01	0.38	0.12	-0.03	0.33
Kurtosis	-1.07	0.09	0.28	-0.12	-0.08

Table 2.7: TVP simulation results when the DGP is deterministic

	mean error	*mean* β	*mean* γ
Mean	1.27	0.37	0.40
Median	1.29	0.37	0.30
Variance	0.04	0.02	0.00
Stdev	0.19	0.15	0.07
Skewness	-0.62	0.25	-0.06
Kurtosis	0.66	1.00	0.99

From the results in Table 2.2, the model with the BP-informed dummy variable performs well. The mean of all the parameters is similar to that of the true parameter values. With the exception of α_1, all of the variances are small. This implies that the number of type I and II errors is relatively low in the simulation. However, α_1, α_2 and γ have high kurtosis which indicate that outliers are large, further away from the mean and occur more frequently than the least squares benchmark.

Evaluating Tables 2.3, 2.4 and 2.5, which report the results of the model with CUSUM, MOSUM and, F-test informed dummies respectively, it is clear that these tests have poor performance. The mean of all of the parameters are far away from the true parameter values. This, in combination with the low variance, implies that the tests consistently misdate the break dates. The poor

performance of CUSUM, MOSUM and F-test is not all that surprising when considering the mechanics of these tests and the type of break that I are dealing with. All of these tests are residual based where a regression as in equation 2.46 *without the dummy variable* is estimated as a first step. The resulting residuals from the regression is then used in the testing procedure. When fitting a model without a dummy variable when the DGP is as in equation 2.46, the resulting residuals will contain only two outliers. Appendix B gives a more detailed explanation and example of why this is the case. Due to the residuals containing only two outliers any cumulative sums thereof will not necessarily be large enough to lie outside of the critical regions. From the CUSUM, MOSUM and F-statistic summary statistics I can see that the test statistics often do not detect any breaks. This can be seen since the means of the parameters are close to those that would be expected from a least squares model that does not control for the structural shift, i.e. model 2.46 without the dummy variable.

From Table 2.6, which reports the MS model results, it is clear that the model performs relatively well. The mean of the parameters for the MS model is similar to that of the least squares benchmark. However, the α_2 parameter has a relatively high variance which indicates that the parameter estimate of α_2 has large deviations from the mean. The small kurtosis (which is less than 1) implies that there are no outliers which are more than one standard deviation away from the mean.

Considering the TVP model results in Table 2.7, the model performs well in some respects. The mean error is relatively small but highly consistent. This can be seen from the low variance and kurtosis. The implication is that the model will always make a positive error in the estimation of the mean compared with the true mean of the DGP. Furthermore, the model always slightly underestimates β and γ.

Comparing the summary statistics of the simulation results, the BP and MS results are numerically relatively similar in the mean. However, the MS model has much higher variance in α_2. The higher variance is a result of the difference in estimated values that are obtained for various starting values of the optimization methods. Section 2.6.2 discusses the effects of optimization procedures on MS models in greater detail. The CUSUM, MOSUM and F-statistic show poor results compared to the BP and MS approaches.

Figure 2.6 shows the box-plot of the estimated error in mean for each of the respective approaches and the LS benchmark.

From Figure 2.6 , the BP, MS and TVP models perform relatively well. The BP shows to have an error distribution that is graphically similar to the LS benchmark. The MS has good performance with respect to its mean error, however, it has much greater variance. The TVP model has the lowest variance out of any of the approaches and is even smaller than the LS benchmark. However, the TVP model consistently overestimates the effects of the structural break in the DGP.

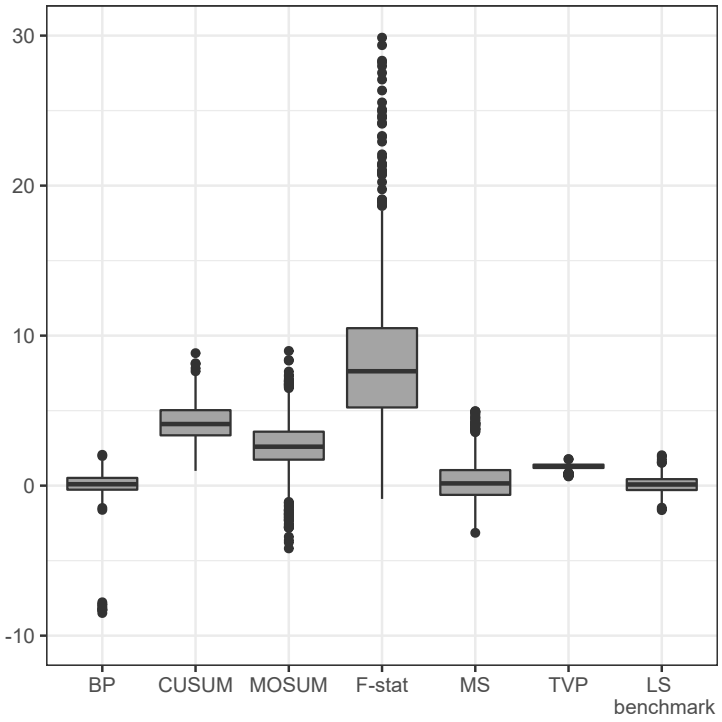


Figure 2.6: Bias for deterministic DGP

Table 2.8 reports the results for the K-S test statistic between the bias of the various approaches and the benchmark LS model. Note that lower values of the K-S statistic implies that the difference between distributions is smaller while larger values implies larger differences. From the results it is clear that the BP approach outperforms the other methods. While the MS model is statistically significantly different from the LS benchmark, the difference between the MS and LS benchmark as well as BP test is much smaller than the other approaches.

Table 2.8: K-S test results for deterministic DGP

	LS benchmark	BP	CUSUM	MOSUM	F-stat	MS
BP	0.05					
CUSUM	0.99***	0.99***				
MOSUM	0.82***	0.81***	0.45***			
F-stat	0.98***	0.97***	0.58***	0.7***		
MS	0.22***	0.19***	0.86***	0.63***	0.98***	
TVP	0.89***	0.87***	0.97***	0.76***	0.97***	0.52***

The result of the RE compared with the LS benchmark is reported in Table 2.9. The relative efficiency compares the MSE between the specified approach

and the LS benchmark. It should be recalled that when $RE < 1$ the approach is more efficient than the LS benchmark while $RE > 1$ implies the approach is less efficient. From the results, all the approaches are less efficient than the LS benchmark. The result is not surprising given the discussion around expected Type I errors for the various structural break tests. Consistent with the previous results the BP, MS and TVP models shows the best performance by a large margin.

Table 2.9: Relative efficiency for deterministic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
3.04	63.91	32.09	926.02	6.79	5.34

2.5.2 Case 2 - Probabilistic shift in DGP

For the second case I consider the case where the DGP is from an MS model. That is when the DGP is specified as:

$$p_t = \begin{cases} 50 + 0.5x_t + 0.5p_{t-1} + \epsilon_t \text{ for } S_t = 1 \\ 50 + 10 + 0.5x_t + 0.5p_{t-1} + \epsilon_t \text{ for } S_t = 2 \end{cases} \tag{2.47}$$

where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$ and the switches between states evolve according to a first order Markov chain with the following probabilities:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \tag{2.48}$$

While the structural changes in this simulation are driven by a probability process, I still can observe the regime realizations in the simulation. This allows us to construct the dummy variable for the LS benchmark according to the regime realizations for each of the simulations. For the results in this section it is important to note that the regime realizations can differ significantly in the frequency of changes and length. This is due to the Markov chain probability process driving the changes. If the regime realizations were kept fixed between each of the simulation iterations the simulated DGP will lose the Markov property.

The summary statistics for the MC simulation results are reported in Tables 2.10 to 2.16.

Table 2.10: LS simulation benchmark when the DGP is probabilistic

	α_1	α_2	β	γ
Mean	50.41	10.04	0.50	0.50
Median	50.34	10.04	0.50	0.50
Variance	5.62	0.17	0.01	0.00
Stdev	2.37	0.42	0.09	0.02
Skewness	0.58	0.04	0.22	-0.60
Kurtosis	1.85	1.25	0.09	1.96

Table 2.11: BP simulation results when the DGP is probabilistic

	α_1	α_2	β	γ
Mean	19.39	0.23	0.40	0.82
Median	13.81	0.20	0.42	0.87
Variance	250.92	29.86	0.08	0.02
Stdev	15.84	5.46	0.29	0.14
Skewness	1.23	-0.03	-0.01	-1.18
Kurtosis	0.73	-0.82	0.76	0.45

Table 2.12: CUSUM simulation results when the DGP is probabilistic

	α_1	α_2	β	γ
Mean	7.26	-0.07	0.39	0.93
Median	6.10	-0.26	0.40	0.94
Variance	36.95	2.00	0.08	0.00
Stdev	6.08	1.41	0.29	0.06
Skewness	5.02	-0.39	-0.25	-5.02
Kurtosis	34.41	0.56	0.22	33.68

Table 2.13: MOSUM simulation results when the DGP is probabilistic

	α_1	α_2	β	γ
Mean	7.03	0.02	0.39	0.94
Median	5.91	0.12	0.39	0.95
Variance	34.87	9.40	0.09	0.00
Stdev	5.91	3.07	0.31	0.05
Skewness	4.94	0.07	0.08	-5.02
Kurtosis	33.86	5.16	0.29	34.65

Table 2.14: F-test simulation results when the DGP is probabilistic

	α_1	α_2	β	γ
Mean	8.24	-0.07	0.37	0.93
Median	6.91	-0.07	0.37	0.94
Variance	46.32	0.98	0.10	0.00
Stdev	6.81	0.99	0.32	0.06
Skewness	4.52	0.10	0.00	-4.54
Kurtosis	29.32	0.51	0.32	29.54

Table 2.15: MS simulation results when the DGP is probabilistic

	α_1	α_2	β	γ	α_2 - α_1
Mean	50.16	60.16	0.13	0.50	9.99
Median	50.13	60.19	0.26	0.50	10.01
Variance	8.22	12.83	0.54	0.00	18.77
Stdev	2.87	3.58	0.74	0.03	4.33
Skewness	0.18	-0.17	-0.39	0.00	0.05
Kurtosis	0.33	0.05	-0.62	0.55	-0.37

Table 2.16: TVP simulation results when the DGP is probabilisitc

	mean error	*mean* β	*mean* γ
Mean	8.99	0.50	0.50
Median	9.01	0.50	0.50
Variance	0.23	0.01	0.00
Stdev	0.48	0.09	0.03
Skewness	-0.31	0.01	-0.58
Kurtosis	-0.15	-0.12	3.07

From the summary statistics, the MS model is the only model that performs well. As noted previously, since the changes in regime is driven by a probability process, I cannot control where and how they occur. The resulting generated process, therefore, often has multiple breaks and every now and then switches in and out quickly. This causes the BP test to miss some breaks and to sometimes pick up the incorrect break dates. As explained in section 2.4.2.1, the BP test relies on a trimming parameter which is the minimum distance between breaks. Therefore, when breaks are too close to one another the test will often miss one of the breaks. This result shows that when there is recurrent collusion, where there are multiple shifts close to one another, the BP framework can produce inconsistent results. The poor performance of the BP test, in this case, is congruent with the results in Antoshin *et al.* (2008)

where the authors show that the BP test often miss breaks when the model enters and exits the same regime more than once.

Figure 2.7 shows the bias of each of the approaches. The bias graphic depicts the same result as the summary statistic. All the other approaches have relatively high positive bias while the BP approach has large negative bias. The bias in the BP result can be ascribed to the fact that the structural break test do find breaks in the simulated DGP but does not detect all the breaks and dates some of them incorrectly. The negative result for the BP is consistent with the theoretical proofs in Boswijk *et al.* (2019); that is that misdating will generally lead to underestimation of the true overcharge.

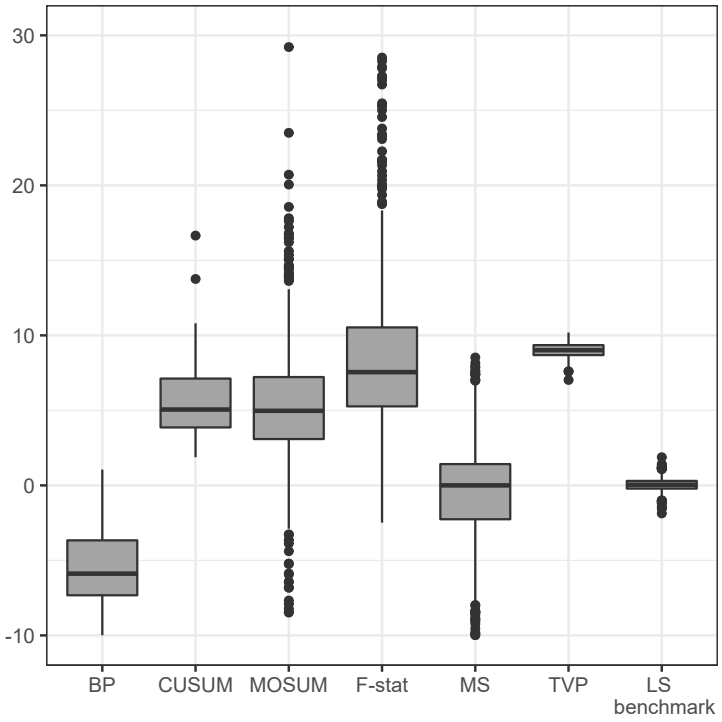


Figure 2.7: Bias comparison for probabilistic DGP

The K-S statistics are reported in Table 2.17. In this case, none of the results produce similar error distributions to the LS benchmark. However, the MS model has the smallest difference. The difference between the distributions can mostly be attributed to the MS model having greater variance between simulation iterations.

Table 2.17: K-S test results for probabilistic DGP

	LS benchmark	BP	CUSUM	MOSUM	F-stat	MS
BP	0.92***					
CUSUM	1***	1***				
MOSUM	0.9***	0.92***	0.17***			
F-stat	0.98***	0.93***	0.35***	0.32***		
MS	0.32***	0.7***	0.79***	0.65***	0.78***	
TVP	1***	1***	0.82***	0.77***	0.52***	0.94***

Results for the measure of relative efficiency is reported in Table 2.18. Considering the RE, all the methods are much less efficient than the LS benchmark. The MS model is the least inefficient but still shows some inefficiencies. This is due to the variation in the estimation results when estimating MS models.

Table 2.18: Relative efficiency for a probabilistic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
714.66	255.04	720.76	2455.45	64.24	463.10

2.5.3 Case 3 - Recurrent shifts in the DGP

For this case the DGP has two shifts in the mean where the process enters the collusive period twice. In this case simulate the following DGP

$$p_t = 50 + 10D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{2.49}$$

with $\epsilon_t \sim IN(0,1)$ and $D_t = 1$ for $t = 11, ..., 30, 41, ..., 60$ and 0 elsewhere.

The summary statistics are reported in Tables 2.19 to 2.25.

Table 2.19: LS simulation benchmark when the DGP is recurrent

	α_1	α_2	β	γ
Mean	50.36	10.06	0.50	0.50
Median	50.31	10.06	0.50	0.50
Variance	4.52	0.16	0.01	0.00
Stdev	2.13	0.40	0.09	0.02
Skewness	0.10	0.07	-0.05	-0.09
Kurtosis	-0.31	-0.09	0.04	-0.31

Table 2.20: BP simulation results when the DGP is recurrent

	α_1	α_2	β	γ
Mean	50.16	10.03	0.50	0.50
Median	50.14	10.01	0.50	0.50
Variance	5.40	0.19	0.01	0.00
Stdev	2.32	0.44	0.09	0.02
Skewness	-0.07	-0.02	-0.13	0.08
Kurtosis	0.73	0.65	0.31	0.79

Table 2.21: CUSUM simulation results when the DGP is recurrent

	α_1	α_2	β	γ
Mean	4.61	0.99	0.38	0.96
Median	4.50	1.01	0.37	0.96
Variance	1.21	1.01	0.09	0.00
Stdev	1.10	1.01	0.30	0.01
Skewness	0.94	-0.39	-0.01	-0.87
Kurtosis	4.35	-0.58	-0.06	3.94

Table 2.22: MOSUM simulation results when the DGP is recurrent

	α_1	α_2	β	γ
Mean	4.53	0.15	0.36	0.96
Median	4.45	0.18	0.35	0.96
Variance	1.08	1.74	0.09	0.00
Stdev	1.04	1.32	0.29	0.01
Skewness	0.61	-1.57	0.00	-0.62
Kurtosis	2.51	15.97	0.04	2.57

Table 2.23: F-test simulation results when the DGP is recurrent

	α_1	α_2	β	γ
Mean	6.31	0.82	0.39	0.94
Median	6.12	0.79	0.39	0.94
Variance	1.54	0.10	0.09	0.00
Stdev	1.24	0.31	0.31	0.01
Skewness	0.69	0.64	-0.01	-0.63
Kurtosis	1.09	2.69	-0.02	0.94

Table 2.24: MS simulation results when the DGP is recurrent

	α_1	α_2	β	γ	α_2 - α_1
Mean	49.82	60.51	0.45	0.50	10.69
Median	49.79	60.51	0.39	0.50	10.58
Variance	4.64	4.69	0.53	0.00	4.30
Stdev	2.16	2.17	0.73	0.02	2.07
Skewness	0.02	-0.13	1.06	0.11	0.20
Kurtosis	-0.72	-0.56	1.48	-0.75	-0.86

Table 2.25: TVP simulation results when the DGP is recurrent

	mean error	*mean* β	*mean* γ
Mean	5.44	0.41	0.68
Median	8.20	0.41	0.90
Variance	9.40	0.06	0.08
Stdev	3.07	0.25	0.29
Skewness	-0.13	0.08	-0.20
Kurtosis	-1.98	0.70	-1.83

Interestingly, the results show that similar to the deterministic case, the BP approach performs reasonably well. The explanation for the discrepancy in results between the recurrent and probabilistic cases for the BP procedure stems from the fact that the regimes are evenly spaced and as a result the breaks are no longer misdated due to the trimming parameter of the BP process. The BP results also have a low variance, similar to that of the LS benchmark. Similar to the results in the previous sections, I find that the other structural break test results have rather poor performance. The MS model again performs reasonably well and has a smaller variance than in the standard deterministic case.

Figure 2.8 shows the box-plot of the error distribution for each of the approaches when estimating overcharge. Once again the BP and LS results are fairly similar with the MS result providing the best alternative but have a higher variance associated to the error possibilities. The MS model performs relatively well in the recurrent case, with a smaller error variance than compared with the deterministic case. Additionally, the error variance implies that the model tends to underestimate the overcharge. The TVP model again produces results with very little variance but constantly overestimates.

The K-S test results are reported in Table 2.26. The BP has an error distribution that is close to the LS benchmark but still statistically significantly different. Once again the MS model has an error distribution that is statistically significantly different from the LS benchmark and BP approach but the

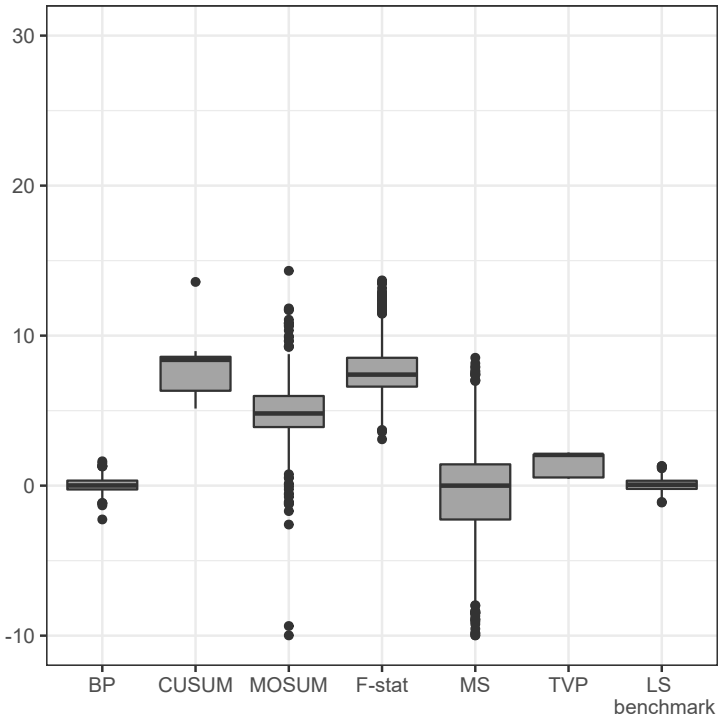


Figure 2.8: Bias comparison for recurrent DGP

difference is much smaller than for the other approaches. This is corroborated by the RE's shown in table 2.27.

Table 2.26: K-S test results for deterministic DGP

	LS benchmark	BP	CUSUM	MOSUM	F-stat	MS
BP	0.06*					
CUSUM	1***	1***				
MOSUM	0.98***	0.97***	0.65***			
F-stat	1***	1***	0.31***	0.64***		
MS	0.32***	0.34***	0.86***	0.63***	0.91***	
TVP	0.85***	0.84***	1***	1***	1***	0.69***

The RE results show that the BP approach is quite similar in this case to the LS benchmark. While the TVP model shows the second lowest RE, it is important to consider this result in context with the other performance metrics presented above. The lower RE for the TVP is due to the low variance of the overcharge errors.

Table 2.27: Relative efficiency for a recurrent DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
1.15	419.89	165.52	363.77	67.23	14.54

2.5.4 Case 4 - Shifts with transition phases

This case represents the DGP that contains transition phases when moving in and out of regimes. For this case the DGP takes the following form

$$p_t = 50 + 10D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t, \tag{2.50}$$

with $\epsilon_t \sim IN(0,1)$ and D_t is encoded as follow:

$$\begin{aligned}
&D_t = 0 \text{ for } t = 1, ..., 20, 80, ..., 100 \\
&D_t = 0.1, 0.2, ..., 0.9 \text{ for each } t \text{ in } [21, 22, ..., 29] \\
&D_t = 1 \text{ for } t = 30, ..., 70 \\
&D_t = 0.9, 0.8, ..., 0.1 \text{ for each } t \text{ in } [71, 72, ..., 79]
\end{aligned}$$

The summary statistics for the MC simulations of each of the approaches is reported in Tables 2.28 to 2.33.

Table 2.28: LS simulation benchmark when the DGP contains a transition phase

	α_1	α_2	β	γ
Mean	50.36	10.06	0.50	0.50
Median	50.31	10.06	0.50	0.50
Variance	4.52	0.16	0.01	0.00
Stdev	2.13	0.40	0.09	0.02
Skewness	0.10	0.07	-0.05	-0.09
Kurtosis	-0.31	-0.09	0.04	-0.31

Table 2.29: BP simulation results when the DGP contains a transition phase

	α_1	α_2	β	γ
Mean	14.31	2.77	0.40	0.86
Median	16.59	3.36	0.40	0.83
Variance	56.45	2.43	0.02	0.01
Stdev	7.51	1.56	0.14	0.08
Skewness	-0.67	-0.86	-0.12	0.68
Kurtosis	-0.78	-0.63	0.28	-0.79

Table 2.30: CUSUM simulation results when the DGP has a transition phase

	α_1	α_2	β	γ
Mean	3.19	0.63	0.35	0.97
Median	2.65	0.54	0.36	0.98
Variance	5.62	0.24	0.03	0.00
Stdev	2.37	0.49	0.17	0.02
Skewness	2.56	1.35	-0.01	-2.66
Kurtosis	9.39	4.01	-0.22	9.88

Table 2.31: MOSUM simulation results when the DGP has a transition phase

	alpha1	alpha2	beta	gamma
Mean	1.69	-0.06	0.37	0.98
Median	1.63	-0.17	0.37	0.99
Variance	0.96	2.35	0.03	0.00
Stdev	0.98	1.53	0.16	0.01
Skewness	0.38	0.03	0.06	-0.38
Kurtosis	1.42	0.89	-0.10	1.44

Table 2.32: F-test simulation results when the DGP has a transition phase

	α_1	α_2	β	γ
Mean	3.83	1.09	0.39	0.96
Median	3.81	1.07	0.40	0.96
Variance	3.55	0.13	0.02	0.00
Stdev	1.89	0.36	0.15	0.02
Skewness	-0.03	-0.01	-0.16	0.01
Kurtosis	-0.05	0.90	-0.16	-0.13

Table 2.33: MS simulation results when the DGP contains a transition phase

	α_1	α_2	β	γ	α_2 - α_1
Mean	49.97	60.48	3.00	0.50	11.54
Median	50.42	59.61	1.55	0.50	11.47
Variance	2.19	4.54	41.41	0.01	12.29
Stdev	1.48	2.13	6.44	0.11	3.51
Skewness	-1.17	1.90	-0.14	-0.12	0.21
Kurtosis	1.63	3.15	1.76	-0.26	-0.64

Table 2.34: TVP simulation results when the DGP has a transition phase

	mean error	*mean* β	*mean* γ
Mean	0.91	0.44	0.61
Median	0.91	0.46	0.50
Variance	0.00	0.03	0.06
Stdev	0.06	0.16	0.25
Skewness	0.15	-0.52	0.65
Kurtosis	-0.11	0.39	-1.24

When comparing the results, the BP procedure has rather poor performance. Upon closer investigation it is no surprise that the test misdates the start and end period of the dates. Due to the misdating the break test not only misses the transition phase but also captures periods that are not truly representative of the change. In many of the simulations the BP would indicate only a single break point or even in some instance no break points at all. This would suggest that the test is rather poorly suited towards detecting changes in regimes where the change takes some time to phase in. The MS process in this instance performs relatively well although there is still some variance in the results.

Figure 2.9 represents the box-plot of the error when estimating the overcharge. Similar to when the DGP was probabilistic the BP approach has relatively poor performance. Again, it underestimates the overcharge due to incorrectly dating the periods of the structural change. The MS model performs relatively well but still has a fair amount of variance in the error. For this particular DGP the TVP model provides the best performance in terms of mean error but again overestimates the size of the break. While there is a small amount of overestimation in the TVP model, it does provide the best result for this case. As mentioned earlier, the estimation results of the TVP model is highly consistent which makes it a favourable approach for this case.

From the K-S results in Table 2.35 none of the error distributions are similar to that of the LS benchmark. In terms of the size of the difference, the MS and MOSUM are the most similar. While the distributions might be similar, the MOSUM has a much higher mean in the error and is still not recommended as an approach for the dating of structural breaks when there is a transition phase.

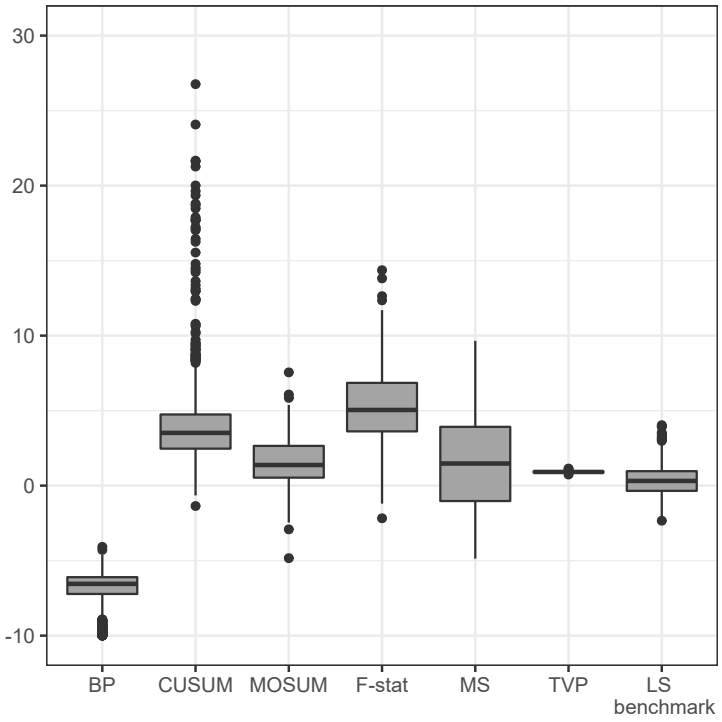


Figure 2.9: Bias comparison for DGP with transition phase

Table 2.35: K-S test results for transition phase DGP

	LS benchmark	BP	CUSUM	MOSUM	F-stat	MS
BP	1					
CUSUM	0.82***	1***				
MOSUM	0.4***	0.99***	0.48***			
F-stat	0.87***	1***	0.32***	0.62***		
MS	0.38***	0.99***	0.18***	0.51***	0.91***	
TVP	0.69***	1***	98***	0.54***	0.97***	0.53***

The RE efficiency results in Table 2.36 show that the MS and TVP models perform the best compared with the LS benchmark. Interestingly, this is the only case in which a model is more efficient than the benchmark. The MS model for this particular DGP also performs relatively better than for the other DGPs.

Table 2.36: Relative efficiency for a transition phase DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
51.14	25.85	5.59	30.07	13.63	0.78

2.6 Empirical simulation challenges associated with Markov-switching models

This section describes some of the simulation complications associated with MS models in the MC simulations. The aim of this section is to assist future researchers in overcoming some practical issues when setting up MC simulation studies for MS models. There are some interesting phenomena that can occur, which I leave for future research to investigate.

In subsection 2.6.1 I discuss the different types of MS models with discussion around software considerations. I show how unknowingly using the incorrect model can lead to inconsistent results. Furthermore, when choosing between the model types there are some clear software limitations that are not well documented. I provide a brief discussion of three main econometric software and discuss the limitations of each. Subsection 2.6.2 explains complications around the numerical optimizers that MS models rely on for estimation. I give an example of how, for certain parameter values, the maximum likelihood function can contain multiple local minima and as a result converge to the incorrect minimum. I provide anecdotal evidence of how these phenomena appear to be related to the autoregressive parameter. In section 2.6.3 I explain how the difficulties with the numerical optimization procedure can translate to poor MC performance and how not accounting for these phenomena can lead to biased results.

2.6.1 Specification and software considerations

There are two main MS specifications, namely, Markov-switching autoregressions (MSAR) and Markov-switching dynamic regression (MSDR), that need to be considered before estimation. There is no direct statistical test to determine which model specification would be most appropriate. Instead, the modeller will have to rely on prior knowledge of the DGP, model fit, and model diagnostics. In the following subsection I explain the difference between the two.

For simplicity, consider a simple autoregressive Markov-switching process with a change in intercept:

$$y_t = c_{s_t} + \rho y_{t-1} + \epsilon_t \tag{2.51}$$

where s_t indicates the regime dependence of the intercept. The MSDR and MSAR estimates switching of the following form:

$$\text{MSDR} : y_t = c_{s_t} + \rho y_{t-1} + \epsilon_t \tag{2.52}$$

$$\text{MSAR} : y_t = c_{s_t} + \rho(y_{t-1} - c_{s_{t-1}}) + \epsilon_t \tag{2.53}$$

The MSDR model is also referred to as a regime-switching intercept model (Krolzig, 2013). This is because the realization of y_t is only dependent on the regime-switching intercept. Hence, there are only two possible intercepts at any given time when there are two regimes. By contrast, in the MSAR specification, the realizations of y_t is dependent on the value of the intercept in the current state as well as the past state. The MSAR is also referred to a Markov-switching mean (Krolzig, 2013). From equation 2.53 it can be seen that if there are two regimes then at any given time there are four possible values of the intercept; $c_1 - \rho c_1$, $c_1 - \rho c_2$, $c_2 - \rho c_1$ and $c_2 - \rho c_2$.

To better understand the dynamics it is easier to write equation 2.52 and 2.53 in MA(∞) form:

$$\text{MSDR} : y_t = \sum_{i=0}^{\infty} \rho^i c_{s_{t-i}} + \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i} \tag{2.54}$$

$$\text{MSAR} : y_t = \rho c_{s_t} + \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i} \tag{2.55}$$

Therefore, in the MSDR model, the effect of a change in regime is averaged out over all past regimes. However, in the MSAR model, the effect of a one-time change is the same for all time periods.

The literature does not typically make the distinction between the two specifications clear. For example, Hamilton (1994), start on page 677 by explaining the MSAR model from Hamilton (1989). Then on page 690 the notation suddenly changes to that of an MSDR without explanation. It is also not helpful that a large amount of literature cites Hamilton (1989), which is an MSAR model, and then proceed to discuss the MSDR variant without explicitly stating that these are different specifications that will not lead to the same results, see for example, Breunig *et al.* (2003) and Hamilton (2010).

When considering software to implement the estimation it is important to note that there are different levels of customization for each of the suites. The optimal software choice depends on the user requirements and technical ability. I investigated three software packages, namely, Eviews, STATA, and R. Both STATA and R[5] can estimate the MSAR and MSDR variants while Eviews can only estimate the MSAR model. STATA also has the advantage of allowing the user to specify starting values for the optimization procedure which – as discussed in subsections 2.6.2 and 2.6.3 – can be pivotal for simulation studies. Eviews again has the advantage of allowing the selection of multiple different optimizers. On the other hand, R encompasses all the functionality from Eviews and STATA, although, much of this needs to be added by the user. I do caution, however, that implementation in R requires programmatic

[5]The MSwM package in R allows for the estimation of MSDR models (Sanchez-Espigares and Lopez-Moreno, 2018). To estimate the MSAR variant, I rely on my own code written in base R.

and extensive statistical knowledge, while Eviews and STATA are more user friendly.

The choice in software is important since there are clear limitations between the packages concerning the MSAR and MSDR specifications. When the DGP is of the MSDR form and an MSAR fit is applied, the model misspecification leads to substantial bias in the parameter estimates in R, Eviews and STATA. In the case of my DGP simulations, the correct specification would be the MSDR model, since changes in the mean are abrupt and the change does not depend on past values of the autoregressive term. When correctly running the MSDR specification on the simulated DGPs, the bias in the parameters was no longer present.

To illustrate the difference in result when choosing between MSAR and MSDR, I generated data with the following DGP:

$$p_t = 50 + 10D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{2.56}$$

with $\epsilon_t \sim IN(0, 1)$. I then fitted an MSAR model with a change in intercept. Figure 2.10 represents the distribution of the autoregressive parameter γ which is clearly biased towards 1.

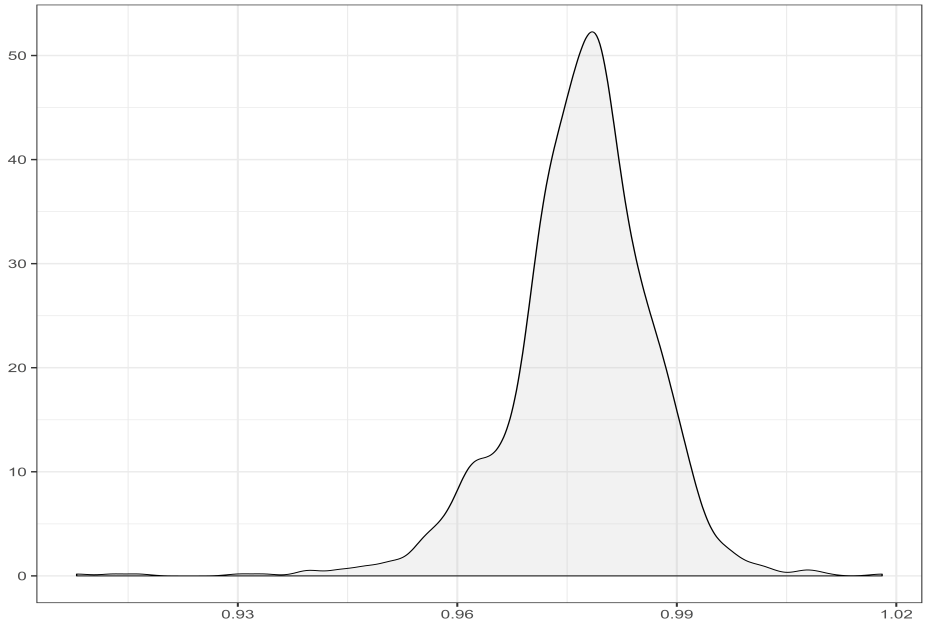


Figure 2.10: Biased AR parameter when fitting MS with change in intercept when the change was in the mean

This fit and biased AR parameter is similar to the LS case where the specification does not control for a break in the intercept (see Appendix B). Similarly, the model provides a relatively good fit but causes the autoregressive

parameter to be biased, which of course translates to bias in the other parameters as well. However, a rather interesting phenomenon occurs in this case. Despite the biased autoregressive parameter, the difference in means between regimes is consistent with what I simulated. Since I set $\alpha_1 = 50$, $\alpha_2 = 10$ and $\gamma = 0.5$ I expect a difference of 20 in the mean between the two regimes. Table 2.37 reports the summary statistic for the simulation along with the difference between mean in the two regimes.

Table 2.37: MS simulation result when incorrectly estimating MSAR

	α_1	α_2	β	γ	difference between means
Mean	8.12	9.74	0.07	0.92	19.77
Median	2.72	3.21	0.09	0.97	20.11
Variance	216.24	312.43	0.11	0.02	7.38
Stdev	14.70	17.68	0.33	0.15	2.72
Skewness	2.27	2.27	-0.14	-2.27	0.03
Kurtosis	3.21	3.21	-0.30	3.22	-1.19

From Table 2.37, it is clear that there is high variance in the parameter estimates within the MC simulation. Interestingly, the difference between the means of the regimes is close to the simulated value of 20 and does not exhibit a high variance. This would suggest that even though the parameters are clearly not what is expected, the MS model still correctly identifies the two regimes and provides a relatively accurate estimate for the mean shift. This result is further corroborated by Table 2.38 where I follow the method proposed in Boshoff and Van Jaarsveld (2019). First I estimated the MS model and save the smoothed regime probabilities. Then I used the probabilities as a dummy-type variable in a subsequent LS estimation. Table 2.38 reports the summary statistics for this type of regression.

Table 2.38: Using MS probabilities as a dummy variable

	α_1	α_2	β	γ
Mean	50.12	10.03	0.51	0.50
Median	50.10	10.00	0.51	0.50
Variance	4.75	0.21	0.01	0.00
Stdev	2.18	0.46	0.10	0.02
Skewness	0.03	0.06	-0.12	-0.03
Kurtosis	-0.51	-0.25	1.18	-0.49

When using the MS smoothed regime probabilities as a dummy-type variable I obtain results that are robust and similar to the LS benchmark (see Table 2.1). The implication of these results is that even when incorrectly estimating a MSAR model when the DGP is of the MSDR form, the smoothed

regime probabilities will still correctly identify the periods of the structural change.

The results in this section are not just dependent on the software as it was replicated in STATA, Eviews and R. For this reason, it is vital to take into account the difference between the MSAR and MSDR specifications as well as the software packages that allow for the different specifications since this can greatly alter the MC results. In the following section I discuss complications associated with numeric optimizers in MS estimations. This also forms an important role in the software choice since software packages are often not clear on what type of optimization procedure is followed. As previously stated, Eviews and R are the most flexible in this regard.

2.6.2 Optimizer complications for MS models

Numeric optimizers require starting values to initialize the optimization procedure. The choice of starting values can either be user specified or are chosen at random depending on the optimizer and pseudo-random number generator of the software. There is an interesting issue with the optimization procedures of MS models that seems to be consistent across multiple software platforms. It should be recalled that the DGP for the data simulation is given by:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta x_1 + \gamma p_{t-1} + \epsilon_t \tag{2.57}$$

The issue arises when choosing values of γ that fall within the range of $4-6$. It appears that for these values of γ there are local minima which cause the optimization procedure to not always converge to its true value. However, when choosing starting values that are close to the true value of the parameter, the procedure will converge to the correct parameter value. Since the problem disappears when starting values closer to the true parameter value are chosen, it would imply that the issue is similar to the Rosenbrock's valley performance test (Rosenbrock, 1960). A graphical explanation, in this case, is useful (see Figure 2.11).

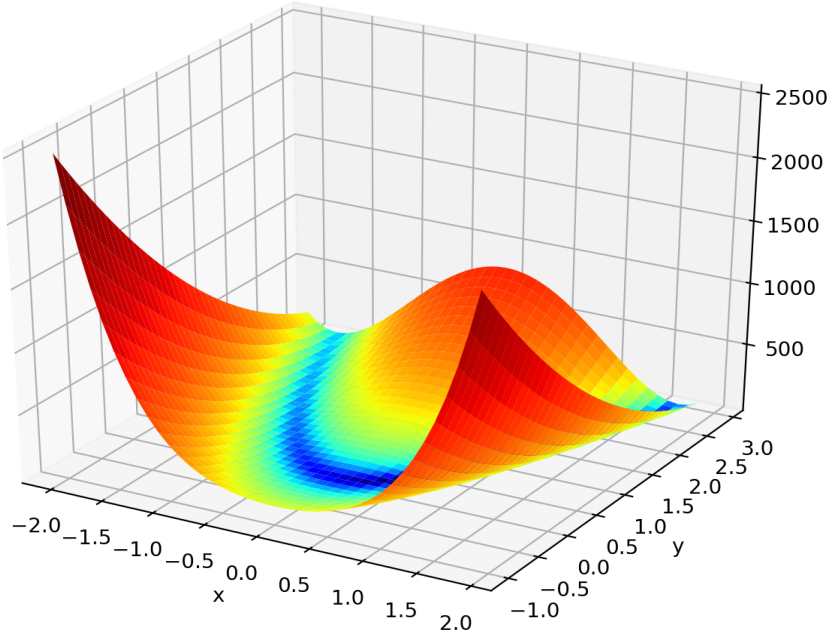


Figure 2.11: Rosenbrock's valley

The graph is generated from the Rosenbrock's function which is given by:

$$f(x, y) = (x - 1)^2 + 100(y - x^2)^2 \tag{2.58}$$

From equation 2.58 it is non-trivial to see that the minimum of the function is reached at $(x, t) = (1, 1)$. Most optimizers will then move rather quickly to the blue valley area in figure 2.11 but will have trouble moving along the valley. The chances of arriving at the global minimum is therefore relatively slim since there are millions of different starting values that will cause the optimization to get stuck in the blue valley area of Figure 2.11.

I have found no literature covering this phenomenon in the context of MS models. This problem is important for both simulation studies and practical applications. For simulation studies, the implication is that one needs to be careful in one's choice of parameter values. Alternatively, starting values close to the true value need to be provided. For practical application, the issue is more complex. To provide starting values that are close to the true estimates is not possible since the true parameter values in applications are unknown.

I recommend that in simulation studies, the authors need to be clear why certain parameter values were chosen and whether it avoids this issue. Alternatively, authors should state that starting values were supplied, and indicate how this influences their results. While many authors, in practical applications, state that their results are consistent for various starting values, it is important to note that convergence to the incorrect minimum is possible for

all their starting values. There are three recommendations that could potentially minimize this issue. First, authors can ensure that different starting values – for robustness checks – are chosen in a way where the different values are far apart. Second, that a large range of different starting values are chosen, and third, that the number of iteration steps without the optimizer is expanded to be as large as possible.

There is a second issue that arises from the use of optimization procedures when using MC simulations. In the MC simulations, some of the iterations will produce outlier parameter results. The reason for this is because the parameter values of the MS process greatly depends on the starting values used by the optimizers of the software. For example, I might find accurate parameter values for 10 different starting values, but inaccurate parameter estimates can be found for one starting value. This result can be caused by a multitude of factors, one of which is the Rosenbrock valley problem described above. If, within the Monte Carlo simulations, the "bad" starting value was used it will report inaccurate parameter estimates for that iteration. These will cause "outlier" results in the Monte Carlo simulations that obviously skew the comparisons. To treat these phenomena, I trim the MC MS simulation results using a z-score to remove possible outliers in the simulation results.

2.6.3 Difficulties with MS MC simulations

As previously mentioned, different starting values of the EM algorithm or numeric optimizers can cause some outliers in the MC simulations. This occurs for the following reason. Suppose that in one of the 5000 repeated simulations starting values were chosen that give inaccurate estimates. These inaccurate estimates will be stored as the parameter values for that run of the MC simulation process. If this occurs multiple times, the end result of the MC simulations will contain multiple outliers. While in 5000 simulations, only a few of the estimates are found to be biased, the bias does influence the overall result. Additionally, when "bad" starting values are chosen that provide inaccurate parameter estimates, I find the magnitude of this inaccuracy to be concerning. Figure 2.12 shows the Monte Carlo simulation result when parameter values for iterations had incorrect estimates due to the optimizer's starting values.

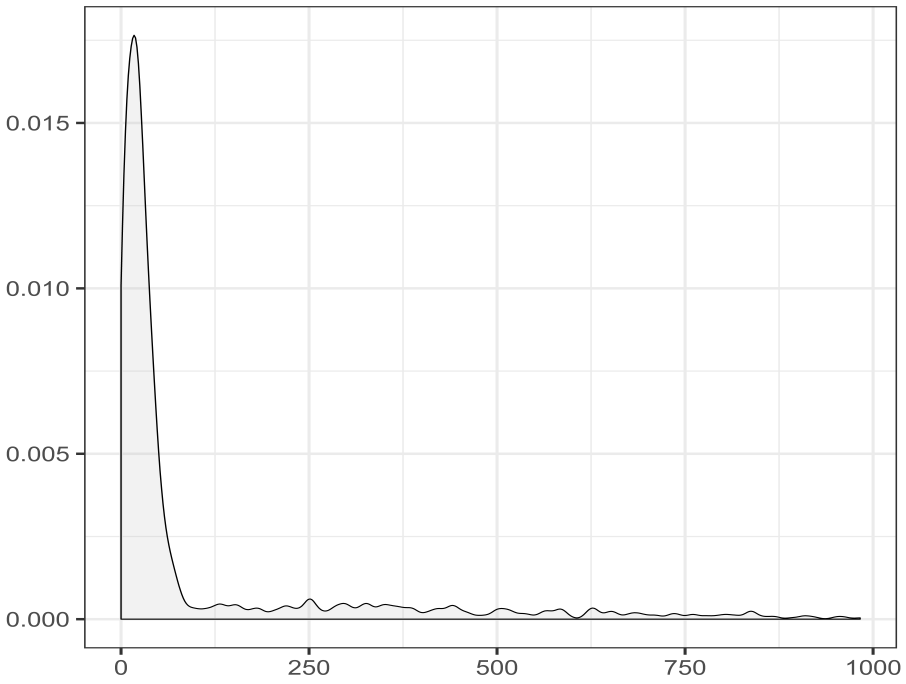


Figure 2.12: Change in mean with outliers in Monte Carlo simulations

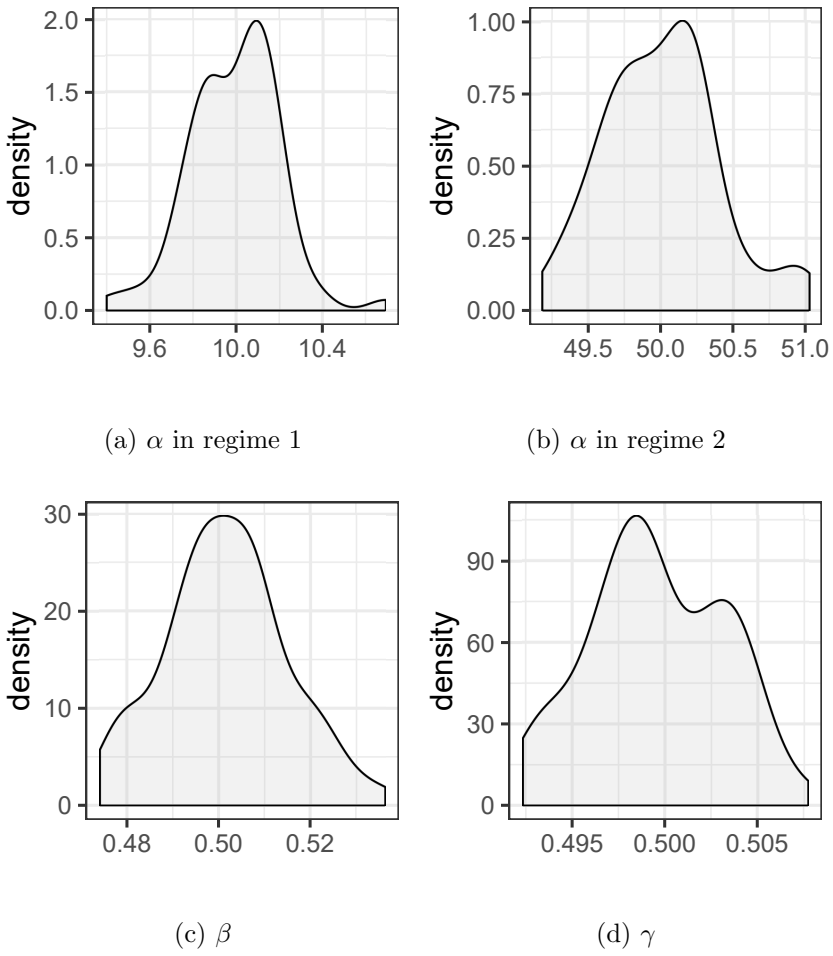
As can be seen from Figure 2.12, the majority of the change-in-mean values are still within a reasonable range – around the *true* value of 20. However, there are enough outliers – which are extremely far from the true change in mean– to influence the overall MC results.

There is limited discussion of this in the Monte Carlo literature. Most studies simply note that outliers are omitted, but do not give a formal explanation as to why or how the omission process is done. For the simulation results reported in this chapter I used a z-score that removes the most significant outliers while retaining 95% of the simulation result.

I suspect that this particular DGP specification causes a Rosenbrock-type problem where there is a local minimum, but the optimization procedure struggles to converge to the global minimum, as described in section 2.6.2. This is due to the fact that when I do not specify starting values that are close to the true parameter values, I obtain two specific sets of results. In 5000 MC simulations, there are two convergence results. Approximately 50% of the simulation results obtain parameter values that are close to the true values while the other 50% obtain parameter values that are rather inconsistent with the simulated DGP. For the incorrect convergence results, the autoregressive parameter γ consistently converges to values around 0.9. The fact that I obtain two sets of results that consistently converge to the same parameter values is

indicative of the Rosenbrock problem, i.e. that there is more than one local minimum and the process often converges to this point.

To illustrate this problem, I create 5000 simulations where I do not specify any starting values using the simple deterministic DGP as specified in equation 2.46. I then split the simulation results into two categories. Category 1 contains the results that are close to the true parameter values specified in the simulation as in equation 2.46. Category 2 contains the parameter estimates that are distinctly not close to the true parameter values. Figure 2.13 illustrates the parameter distributions for category 1 while Figure 2.14 shows the category where the parameter estimates are not close to the true parameter values.



(a) α in regime 1

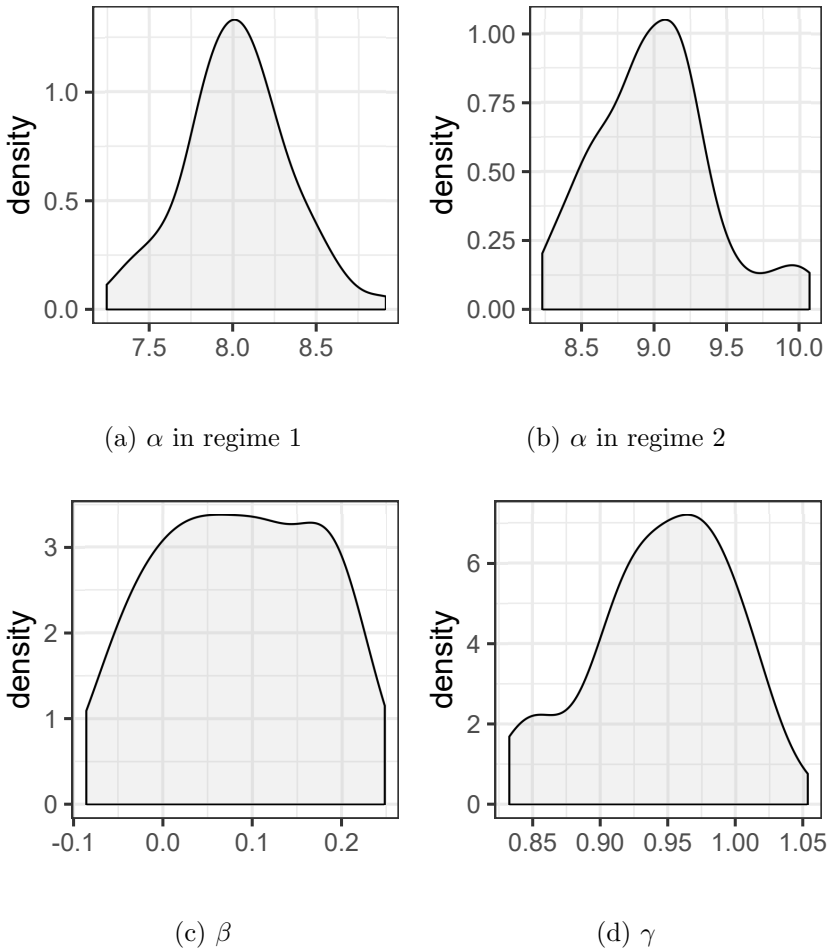
(b) α in regime 2

(c) β

(d) γ

Figure 2.13: Parameter distributions for split simulation results – Category 1

To split the parameter results into the two categories I apply a simple separation rule. All parameters that are within two standard deviations of the mean for a specified standard deviation threshold are included in category 1.



(a) α in regime 1

(b) α in regime 2

(c) β

(d) γ

Figure 2.14: Parameter distributions for split simulation results – Category 2

That is, for α all iterations that fall within $[8, 12]$ for regime 1 and $[48, 52]$ are included in category 1. For β, all observations that fall within $[0.46, 0.54]$ are included in category 1. For γ, all observations that fall within $[0.46, 0.54]$ are included in category 1. It is important to note that for a single iteration in the simulations to be included in category 1, all the parameter estimates need to fall within the above specified ranges. Therefore, an estimation result that is included in category 1 does not mean that only a single parameter was close to the true value, but implies that all the parameter estimates were close to their true values. All estimation results that do not produce parameter estimates that fall within the above specified ranges are included in category 2.

From Figures 2.13 and 2.14 there are two clear results that emerge. There is a significant amount of estimations that converge to their true values. However, when the parameters do not converge to their true values, they still end up converging to the same incorrect values rather consistently. Based on this result, it is likely that more than one local minimum exists for the specification

in this dissertation, as in the Rosenbrock problem.

Some peculiar results are found when making slight alterations to the MS estimations or choice of parameter values. These results and more detailed discussions thereof are presented in Appendix C. The first result is that when data is simulated from a standard deterministic DGP, as in equation 2.46, and an MS model is fitted to the data where I allow for both the intercept and autoregressive parameters to change regimes, the issues related to the Rosenbrock problem and starting values seem to disappear. In this case, the model consistently estimates the correct parameter values and identify the accurate timing of the structural change. The autoregressive parameter produces an interesting result where both the values between regime 1 and regime 2 are still close to the correct value of 0.5. That is, the autoregressive parameter in regime 1 has a mean of 0.48 while for regime 2 it has a mean of 0.52. The second interesting result is when the simulated process has an autoregressive parameter that is higher than 0.5. I find that when this value is higher than approximately 0.6, the simulation complications explained in section 2.6 are no longer a concern. The results for this case is reported in Appendix C.

I suspect that the reason for both of the above-mentioned results, is related to the Rosenbrock problem and optimization procedures used in the estimation. Similar than to those in subsection 2.6.1, the results in Appendix C are not dependent on the software used and were replicated in STATA, Eviews and R. The evidence in appendix C is anecdotal and not yet conclusive. I leave investigation of these phenomena for future research.

Given this experience with parameter selection in MC simulation experiments, I caution researchers to take care when setting up these experiments. Not specifying the starting values and not carefully examining the results of each of the MC iterations can lead to incorrect conclusions. Furthermore, in future research more explicit explanations should be given as to why certain parameter values are chosen, since this choice can greatly alter the results.

2.7 Conclusion

A critical component in overcharge estimation is the correct determination of the effective collusion dates. In principal, this is related to the time series literature of dealing with unknown break dates. Boswijk *et al.* (2019) shows that incorrect dating of the effective cartel dates will lead to an underestimation of the overcharge. To overcome this issue, the applied literature suggest the use of structural break tests and regime-switching models to determine the break dates.

In this chapter, I investigate the performance of various structural break tests and regime-switching models when used in overcharge estimation. This is a slight departure from the typical time series literature that investigate power and size of the various break tests and estimation efficiency of the regime-switching models. I investigate these tests and models from a new angle to see how they perform when used to determine the size of the break in subsequent regressions. In the case of structural break tests we, therefore, investigate how type I and II errors translate to subsequent regressions when structural break tests are used to inform the encoding of dummy variables.

I test these approaches by performing MC simulations and consider their performance on four different possible DGPs. First, I consider the standard case where the structural change is a deterministic change in the intercept. Second, I consider the case where the changes in the DGP is driven by an underlying Markov process. Third, I consider the case where there are multiple deterministic changes, and fourth, the case where the structural change has a transition phase. This chapter also provides a discussion of the empirical issues related to MS models in MC simulation studies in order to assist future researchers.

The results show that in general the BP and MS approaches perform best. The BP approach, however, failed in the probabilistic case and in the case where there was a transition phase in the structural shift. In both of these cases the MS approach provided robust results. In the simple deterministic case and the recurrent case, the BP approach performs the best and produces results that are similar to the benchmark model. While the MS model produces good results for all the cases, there is much more variation in the estimates. This implies that while it is a robust method regardless of the nature of the structural change, greater care needs to be taken when interpreting the results.

The econometric concerns of dealing with structural change is not separable from issues related to unit roots. As a result, much of the research related to structural change have had analogue developments in the unit roots literature. The following chapter considers the interplay between unit roots and structural breaks. Specifically, the chapter considers the effects that unit roots have on dummy variable parameters and cointegration tests.

Chapter 3

Don't trust the dummy: The effects of unit roots on dummy variable coefficients

3.1 Introduction

time series models are often employed to estimate the relationship between observed variables over a specific time frame. A common assumption in these models is that the relationship between the variables remains constant over the entire sample period. However, there are many cases where exogenous factors – such as wars, policy changes, and unexpected events – cause sudden changes in the underlying relationship. In time series analysis, dummy variables are frequently used to isolate or capture the effects of certain periods that are systematically different from others. As a result, dummy variables have become a staple numeric stand-in to control or estimate the effects of a structural change.

Many economic time series variables exhibit non-stationarity. Even so, especially in the literature on cartel overcharges non-explosive economic time series are assumed to be strictly or weakly stationary, implying the series has a constant mean and variance over time. A quick glance at a few graphs of economic time series quickly reveals that stationarity is often violated (see Nelson and Plosser (1982)). A specific form, and one often found in price overcharge regressions, involves unit roots. If a time series has a unit root, it implies an unpredictable trend with particular implications for these models, as explained below.

The seminal work of Granger and Newbold (1974) gave rise to a wealth of literature on how to deal with unit roots in regressions. An important consideration to take into account is the effect that unit roots have on the inference of the regression coefficients. In general, the coefficients will no longer have asymptotically normal distributions and inference can therefore no longer

be drawn using Student's t-tests. A common response when dealing with unit roots is to test for cointegration to ensure that the results are not spurious. Yet even when cointegration can be established, the coefficients in the cointegrating regression will still not have asymptotically normal distributions.

The basic premise of cointegration tests is to establish the existence of common trends and therefore a long-run relationship between the variables in the regression exists. It is therefore not surprising that cointegration tests are sensitive to model specification. If the model is misspecified the long-run relationship would not be correctly identified and cointegration tests will not detect this relationship.

When a dummy variable is included in a cointegrating regression it is meant to capture the structural change within the model. Construction of the dummy variable is typically informed by structural break tests or a priori information about the dates associated with the structural change. When a dummy variable is constructed in this manner it is possible to misdate the break dates and as a result have a misspecified dummy variable. The effect that this would have on cointegration tests is somewhat ambiguous. The dummy variable is binary and therefore does not contain common trends with other variables in the regression. However, misspecifying the dummy variable will influence the residuals and model coefficients that are used in cointegration tests. Therefore, misspecification of the dummy variable can influence cointegration tests through its impact on the residuals and coefficients in the model.

The focus of this chapter is on the effects that unit roots have on inference related to dummy variable coefficients. I am mainly concerned with the effects on dummy variables where the dummy is used to capture the effects of innovative outliers. In other words, I focus on dummy variables that are used to capture a mean shift in the regression, or changes in regime. The chapter also shows that cointegration tests can only solve the dummy variable problem if the break dates are appropriately specified.

Developing a better understanding of the inferential conjectures of dummy variables in cointegrated regressions is important for applied work. Dummy variables are frequently used to estimate the effects of some exogenous qualitative event, such as policy changes. As many economic time series are non-stationary, applied work must account for the effects that unit roots can have on the inference of dummy variables and cointegration tests.

As will be discussed in the literature review (section 3.2) with formal mathematical proofs derived in sections 3.3, the asymptotic results for cointegrated regression coefficients is well understood. However, there has been little research into the related asymptotics of dummy variable coefficients. Although the theory is not explicitly covered in the literature, a general intuition of the theoretical results can easily be obtained by considering the dummy variable as a shifting intercept. There is currently a lack of research into the finite sample distributional properties of dummy variables in cointegrated regressions. The results in this chapter attempt to address this gap.

In keeping with the approach in the preceding chapter, this chapter illustrates the importance of considering the inferential effects in the presence of unit roots, by reference to application from the cartel damage models. As discussed in Chapter 2, an important consideration in the litigation of cartels is estimating overcharge. That is, the difference between the observed market price under collusion and the counter-factual price that would have prevailed in the absence of the illicit conduct. Dummy variables are frequently used in overcharge estimation, and in practice, the significance of the dummy variable is often contested. I therefore extend the simulation results to illustrate inferential problems that unit roots can induce in overcharge estimations.

This chapter aims to make the following contributions. First, it empirically illustrates the difference between the "true" (or empirical) distribution and Student's t-distribution of dummy variable coefficients in finite samples of cointegrated series. This is done to demonstrate the substantial risk of drawing incorrect inferences on the significance of dummy variable coefficients. This chapter also evaluates the related size distortions of t-tests in this context. Second, when performing bootstrapping on nonstationary data, the chapter discusses an important adjustment that should be made when the true data generating process (DGP) contains a dummy variable. Last, I consider the effects that misspecifying the break dates will have on cointegration tests.

The methodological framework consists of simulating various nonstationary cointegrated series and obtaining the difference between the "true" distribution of the dummy coefficient for each of the simulations. I focus on the case where there was a single cartel period and therefore two break dates associated with the start and end of the cartel period. The simulations are repeated for various sample sizes T and dummy variable lengths[1]. The results show that in small samples and for shorter dummy variable lengths the "true" distribution of the dummy coefficient differs substantially from that of a Student's t-distribution. In the overcharge application I show that, even when misspecifying the dummy variable, incorrect inferential conclusions can still be drawn. To show the complications in cointegration testing, I simulate data and estimate subsequent regressions and cointegration tests. For each of the regressions, I deliberately incorrectly encode the dummy variable to span over a period that includes both competitive and collusive conduct (i.e. does not relate to collusive conduct only).

The rest of this chapter is organized as follows. Section 3.2 discusses the related literature. Section 3.3 provides a theoretical discussion of how the asymptotic distributions of regression coefficients differ when considering cointegrated regressions. Section 3.4 shows how the theory extends to a practical application in overcharge estimations. The methodology is set out in section 3.5, with the results reported in section 3.6. Section 4.4 provides guidelines for

[1]I define the length of the dummy variable as the number of consecutive 1s contained in the encoding

practice, and section 3.7 concludes.

3.2 Related literature

Many economic variables contain trends that can be described as stochastic. Stochastic trends are induced by the persistent cumulation of past events and referred to as unit root processes. There are many reasons why economic data may contain such trends. For example, the current level of technology is dependent on the persistence of previous knowledge so that the current level is an accumulation of past innovations. Economic variables depending on any variable that contains a stochastic trend will "inherit" this nonstationarity, and in turn, will transmit this property to other variables. Concurrently, variables that "inherited" this nonstationary property, as a result of their dependence, will trend together. As a result, certain linear combinations can form a stationary system and are therefore said to be cointegrated. The presence of nonstationary data clearly violates the classical linear model assumptions. The econometric literature has therefore given careful consideration to dealing with the effects of this violation. Specifically, the two foremost considerations in applied work where the variables are nonstationary, is the framework of cointegration testing and the subsequent effects on hypothesis testing and coefficient inference.

Yule (1926) was the first to observe that significant correlations are often obtained between nonstationary time series that should definitely be unrelated. Only many years later, Granger and Newbold (1974) emphasized in a regression framework that an apparent 'significant relation', where the residual series are strongly autocorrelated, is a result of nonsense regressions. The intuition behind their result is rather simple. Any two nonstationary series will be drifting either up or down whether they are related or not. If both these series drift in the same direction, performing a regression will find a significant relationship even when there is none. Phillips (1986) provides a technical analysis of the sources and consequences of these nonsense regressions.

To address the concerns of possible nonsense regressions, the concept of cointegration was introduced by Granger (1981), followed by the statistical analysis using regression methods in Engle and Granger (1987). The simple intuition behind cointegration is to test whether the stochastic trends between time series are related. If the stochastic trends are related, it follows that one variable 'inherited' the nonstationary property from a related causal variable. The key point here is that due to the relationship between the stochastic trends, there exists a linear combination that will form a stationary residual. Therefore, the strong autocorrelated residual problem of Granger and Newbold (1974) will be resolved. Building on the work of Granger and Newbold, Johansen (1988), and Pesaran and Shin (1998) extends the cointegration analysis to the vector autoregressive (VAR) and autoregressive distributed lag (ARDL) models respectively.

Cointegration analysis has become an overriding requirement for any economic model using nonstationary data. While cointegration testing solves the problem of spurious regressions, the asymptotic properties of the regression coefficients, statistical tests, and model diagnostics are different from regressions with stationary series. Stock (1987) and Phillips and Durlauf (1986) develop the OLS asymptotic theory for cointegrated systems. The authors show that OLS coefficient estimates are *super consistent* but converge in distribution to a non-normal random variable not necessarily centred around zero. Therefore, the conventional use of t-statistics in determining coefficient significance no longer applies. Park and Phillips (1988) extends and simplifies earlier work by developing an asymptotic theory of sufficient generality to accommodate most regressions in a multivariate framework that permits both deterministic and stochastic trends.

The asymptotic non-normality results of Stock (1987), Phillips and Durlauf (1986) and Park and Phillips (1988) hold for all coefficients in a cointegrated regression model, including the intercept. Schmidt and Phillips (1992), Xiao and Phillips (1999) and Hansen (1995) use Monte Carlo simulations to show that in finite samples, commonly used testing procedures suffer from size distortions. The degree of this distortion is dependent on nuisance coefficients. Yet, surprisingly, there are no analytical studies of how these effects might translate to dummy variable coefficients. It is tempting to conclude that dummy variables, as defined in this chapter, is simply a shifting intercept and should, therefore, have the same asymptotic properties as the intercept. However, the simulation results show that the distributional effects are much more nuanced. The magnitude of the difference between the distribution of the dummy variable coefficient and a t-distribution depends on both the total sample size T and the length of the dummy variable.

Considering dummy variables in the context of cointegrating systems is by no means a new consideration in the literature. Nielsen (2004) shows that misspecified dummies and uncontrolled outliers can distort the inference on cointegration rank in finite samples. This result further motivates the importance of understanding the distributional properties of dummy variables in cointegrated regressions. Misspecified or insignificant dummy variables can lead to falsely rejecting the presence of cointegration. Adequate significance testing of the dummy variable coefficient can reduce the probability of falsely rejecting cointegration.

The simulation results attempt to explain the nonstationary effects on coefficient inference in a different manner than what is found in previous literature. Typically, Monte Carlo simulations or bootstrapping is used to empirically derive asymptotic approximations. While important, these results are difficult to translate into practical applications. I therefore provide an alternative presentation of the results. Through the use of Monte Carlo simulations and bootstrapping, I calculate the difference between the true distribution of the dummy variable coefficient and the Student's t-distribution for various sample

sizes. By illustrating the results in this manner, it is much simpler to see how the distributions differ from that of a t-distribution.

To illustrate an application where the results can guide and improve existing methods I consider the cartel damage literature. In recent years, there has been a steady increase in the number of prosecutions for price-fixing cartels. Consequently, the estimation of damages caused by these cartels has received increasing practical and academic attention. In the Green chapter by the European European Commission (2005), the quantification of damages was identified as one of the key barriers to the promotion of further anti-trust damage actions. In the subsequent White chapter (European Commission, 2008) the European Commission announced their plan to derive a unified economic framework which provides non-binding guidance on the quantification of damages. The latest working chapter of this guidance was published in 2013, (European Commission, 2013).

Econometric models have since become the standard methodology for estimating cartel damages (Connor, 2007, 54). One of the most widely used and studied methods is the dummy variable approach (see section 3.4 for details). Academic and practical applications have not yet concerned themselves with the effects that nonstationary data can have on econometric models. In the application, I shed some light on the effects of this oversight on the dummy variable approach to cartel overcharges.

When using a dummy variable in overcharge estimation, the determination of collusive periods – on which the dummy specification relies – can be sensitive to the chosen collusive periods. The standard models on price overcharge rely on exogenously determined collusive periods, often provided by the court. Ideally, the court can establish the correct collusive periods from hard evidence. Unfortunately, as discussed in Chapter 2, such evidence may not always be available or reliable. Furthermore, legal liability may not always coincide with cartel effects. The dummy variable can therefore easily be misspecified. It is therefore important to understand how misspecification of the dummy variable can influence cointegration tests.

3.3 Effects of unit roots on asymptotic properties of regression coefficients

A source of nonstationarity in economic time series is the existence of trends. The literature distinguishes between two types of trends: deterministic and stochastic. When a deterministic trend is present, the mean of the process is replaced by a linear function of the date (t). Such a process is often referred to as being *trend stationary* since subtracting the trend from the series results in a stationary process. A stochastic trend can be interpreted as allowing a different trend at each point in time. When considering stochastic trends,

nonstationarity is induced as a result of the persistence of past effects that accumulate over time. In this chapter – for reasons that will become apparent in the following explanation – I focus on nonstationary series where the cause of nonstationarity is a result of stochastic trends.

To further explain the effect of nonstationarity on the distributions of the coefficients in OLS, consider the following system of cointegrated variables:

$$y_t = \alpha + \beta x_t + u_t \tag{3.1}$$
$$x_t = x_{t-1} + \varepsilon_t \tag{3.2}$$
$$u_t \sim IN(0, \sigma_u^2), \quad \varepsilon_t \sim IN(0, \sigma_\varepsilon^2), \quad E(u_t \varepsilon_s) = \delta_{ts}\sigma_{u\varepsilon},$$

where δ_{ts} is the Kronecker delta. Through recursive substitution it can be shown that $x_t = x_0 + \sum_{i=1}^{t} \varepsilon_t$. The initial value x_0 will always be present in the process and any shock to x_t will stay in the process forever. Therefore, since past effects accumulate over time, x_t has a stochastic trend. Since y_t depends on x_t it will similarly contain a stochastic trend. Equation 3.1 can be re-written in matrix notation as:

$$y_t = X_t' A + u_t \tag{3.3}$$

with

$$X_t = \begin{bmatrix} 1 \\ x_t \end{bmatrix} \text{ and } A = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

The OLS estimator of A is given by

$$\hat{A} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t y_t \tag{3.4}$$

Substituting equation 3.3 into equation 3.4

$$\hat{A} = \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t \left(X_t' A + u_t \right) \tag{3.5}$$

through multiplication and re-arranging the terms, equation 3.5 can be written as:

$$\begin{aligned} \hat{A} - A = \begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{bmatrix} &= \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t u_t \\ &= \left(\sum_{t=1}^{T} \begin{bmatrix} 1 \\ x_t \end{bmatrix} \begin{bmatrix} 1 & x_t \end{bmatrix} \right)^{-1} \sum_{t=1}^{T} \begin{bmatrix} 1 \\ x_t \end{bmatrix} u_t \\ &= \begin{bmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} {x_t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} x_t u_t \end{bmatrix} \end{aligned} \tag{3.6}$$

To obtain the distribution of $\hat{A}$ as $T \xrightarrow{\infty}$ equation 3.6 is multiplied by the scaling matrix $\mathbf{T}$, where $T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}$. Hence, the distribution of $\hat{\alpha}$ and $\hat{\beta}$ can be written in matrix notation as

$$\begin{bmatrix} T^{1/2}(\hat{\alpha} - \alpha) \\ T(\hat{\beta} - \beta) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\sum_{t=1}^{T} x_t \\ T^{-3/2}\sum_{t=1}^{T} x_t & T^{-2}\sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} u_t \\ T^{-1}\sum_{t=1}^{T} x_t u_t \end{bmatrix} \tag{3.7}$$

In this case, equation 3.7 does not converge to a deterministic limit, but instead converges to a stochastic variable. To derive the limiting distribution of the coefficient in this context, involves the use of Brownian motions. Brownian motions can be approximated by random walks; thus, a Brownian motion can be thought of as a random walk with a large number of steps. This relationship allows the study of the asymptotic properties.

The original derivation of the asymptotic theory can be found in Phillips and Durlauf (1986) and Stock (1987). I provide a simplified explanation and representation of the results below.

From Lemma 3.1 in Phillips and Durlauf (1986) I have that

$$T^{-3/2}\sum_{t=1}^{T} x_t \xrightarrow{d} \sigma_\varepsilon \int_0^1 B_\varepsilon(r)\mathrm{d}r \tag{3.8}$$

$$T^{-2}\sum_{t=1}^{T} x_t^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 B_\varepsilon(r)^2\mathrm{d}r \tag{3.9}$$

$$T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t \xrightarrow{d} \int_0^1 B(r)\mathrm{d}B(r) \tag{3.10}$$

also note that by the properties of u_t it follows that

$$T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{d} \sigma_u B_u(1) \tag{3.11}$$

From equations 3.8, 3.9 and 3.11, I have the limiting distributions of $T^{-3/2}\sum_{t=1}^{T} x_t$, $T^{-2}\sum_{t=1}^{T} x_t^2$ and $T^{-1/2}\sum_{t=1}^{T} u_t$. Therefore, I need only derive the limiting distribution of $T^{-1}\sum_{t=1}^{T} x_t u_t$. To obtain the limiting distribution of $(T^{-1}\sum_{t=1}^{T} x_t u_t)$, the calculations are simplified by first conditioning u_t on ε_t:

$$\begin{aligned} u_t &= \zeta\varepsilon_t + \upsilon_t \\ \zeta &= \sigma_{u\varepsilon}/\sigma_\varepsilon^2 \\ \sigma_\upsilon &= \sigma_u^2 - \sigma_{u\varepsilon}^2/\sigma_\varepsilon^2 \end{aligned} \tag{3.12}$$

It follows that $E(\varepsilon_t \upsilon_s) = 0 \; \forall \; t \neq s$.

By substituting equation 3.12 into $T^{-1}\sum_{t=1}^{T} x_t u_t$ it follows that

$$\begin{aligned}(T^{-1}\sum_{t=1}^{T} x_t u_t) &= T^{-1}\sum_{t=1}^{T} x_t(\zeta\varepsilon_t + \upsilon_t) \\ &= \zeta(T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t) + \zeta(T^{-1}\sum_{t=1}^{T}\varepsilon_t^2) + (T^{-1}\sum_{t=1}^{T} x_{t-1}\upsilon_t) \\ &+ (T^{-1}\sum_{t=1}^{T}\varepsilon_t\upsilon_t)\end{aligned} \tag{3.13}$$

Using equation 3.10 it follows that

$$T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t \xrightarrow{d} (\sigma_\varepsilon^2/2)(B_\varepsilon(1)^2 - 1) \tag{3.14}$$

From the properties of ε_t it follows that

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \sigma_\varepsilon^2 \tag{3.15}$$

From Lemma 3.1 (e) in Phillips and Durlauf (1986) it follows that

$$T^{-1}\sum_{t=1}^{T} x_{t-1}\upsilon_t \xrightarrow{d} \sigma_\varepsilon\sigma_\upsilon\int_0^1 B_\varepsilon(r)\mathrm{d}B_\upsilon(r) \tag{3.16}$$

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t\upsilon_t \xrightarrow{p} 0 \tag{3.17}$$

Equation 3.17 follows from the fact that ε_t and υ_t are independent (by construction) and both processes are identically and independently distributed with zero means and variances σ_ε^2 and σ_υ^2 respectively. By using equations 3.14 to 3.17 the limiting distribution of 3.13 can be deduced which is given by

$$T^{-1}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} (\zeta\sigma_\varepsilon^2/2)(B_\varepsilon(1)^2 + 1) + \sigma_\varepsilon\sigma_\upsilon\int_0^1 B_\varepsilon(r)\mathrm{d}B_\upsilon(r) \tag{3.18}$$

By using the results in equations 3.8, 3.11 and 3.18 the limiting distributions of $\hat{\alpha}$ and $\hat{\beta}$ can be written as

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\beta}-\beta)\end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & \sigma_\varepsilon\int_0^1 B_\varepsilon(r)\mathrm{d}r \\ \sigma_\varepsilon\int_0^1 B_\varepsilon(r)\mathrm{d}r & \sigma_\varepsilon^2\int_0^1 B_\varepsilon(r)^2\mathrm{d}r\end{bmatrix}^{-1} \cdot \begin{bmatrix} \sigma_u B_u(1) \\ (\zeta\sigma_\varepsilon^2/2)(B_\varepsilon(1)^2+1) + \sigma_\varepsilon\sigma_\upsilon\int_0^1 B_\varepsilon(r)\mathrm{d}B_\upsilon(r)\end{bmatrix} \tag{3.19}$$

It can be proved that (see Park and Phillips (1988))

$$\int_0^1 B_\varepsilon(r)\mathrm{d}B_v(r) \xrightarrow{d} N\left(0, \int_0^1 B_\varepsilon(r)^2\mathrm{d}r\right) \tag{3.20}$$

Therefore, by using equation 3.20 equation 3.19 can be expressed as

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\beta}-\beta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & \sigma_\varepsilon \int_0^1 B_\varepsilon(r)\mathrm{d}r \\ \sigma_\varepsilon \int_0^1 B_\varepsilon(r)\mathrm{d}r & \sigma_\varepsilon^2 \int_0^1 B_\varepsilon(r)^2\mathrm{d}r \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sigma_u B_u(1) \\ (\zeta\sigma_\varepsilon^2/2)(B_\varepsilon(1)^2+1) + N\left(0, \int_0^1 B_\varepsilon(r)^2\mathrm{d}r\right) \end{bmatrix} \tag{3.21}$$

The final result in equation 3.21 shows why the t-statistic, of $\hat{\alpha}$ and $\hat{\beta}$, will typically not follow a standard normal distribution. However, if $\zeta = 0$ it is simple to deduce from equation 3.21 that the t-statistic can be reduced to a standard normal distribution. It should be recalled that ζ is effectively the relationship between the error terms of x_t and y_t. Therefore, in order for the t-statistic of $\hat{\alpha}$ and $\hat{\beta}$ to converge to a standard normal distribution, x_t is required to be strictly exogenous[2] for the estimation of β. Relaxing the strict exogeneity assumption can be asymptotically negligible in estimating β due to super consistency (see, Phillips and Hansen, 1990). However, relaxing the assumption implies that $\zeta \neq 0$ and therefore $\hat{\alpha}$ and $\hat{\beta}$ will no longer have a standard normal distribution.

From equation 3.21, when $\zeta = 0$, it follows that[3]

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\beta}-\beta) \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & \sigma_\varepsilon \int_0^1 B_\varepsilon(r)\mathrm{d}r \\ \sigma_\varepsilon \int_0^1 B_\varepsilon(r)\mathrm{d}r & \sigma_\varepsilon^2 \int_0^1 B_\varepsilon(r)^2\mathrm{d}r \end{bmatrix}^{-1} \cdot \begin{bmatrix} N(0,1) \\ N\left(0, \int_0^1 B_\varepsilon(r)^2\mathrm{d}r\right) \end{bmatrix} \tag{3.22}$$

Therefore, given that x_t is strictly exogenous, the distribution of $\hat{\alpha}$ and $\hat{\beta}$ is mixed normal with $\left(\int_0^1 B_\varepsilon(r)^2\right)$ as mixing variate.

In practical applications, however, strict exogeneity will often be violated. For example, including lagged values of the endogenous variable will result in violation of this assumption. The assumption therefore rules out any type of dynamic relationship between the predictors and the residuals. Additionally, results from my Monte Carlo simulations show that – even when I have full

[2]Note that for strict exogeneity the requirement is $E(u_t|X)$. Therefore, x_{sj} must be uncorrelated with u_t $\forall s$, even when $s \neq t$

[3]I use the notation $\Longrightarrow$ to indicate weak convergence.

control over the DGP – strict exogeneity is often violated in smaller samples. Even comparatively weak relationships between x_t and u_t may cause a violation and can have significant effects on the distribution of β. It is therefore important that careful consideration should be given when basing inferences on model coefficients when the data is nonstationary.

Comparing the above to a case where the data is characterised by a deterministic and stochastic trend. Suppose that, instead of equation 3.2, x_t contains a deterministic trend and is generated by:

$$x_t = \psi t + x_{t-1} + \varepsilon_t \tag{3.23}$$

Note that

$$T^{-3/2}\sum_{t=1}^{T} x_t u_t = \psi T^{-3/2}\sum_{t=1}^{T} t u_t + T^{-3/2}\sum_{t=1}^{T} u_t \left(\sum_{s=1}^{t} \varepsilon_s\right) \tag{3.24}$$

From the results in Stock (1987) and from the orders of magnitude involved it follows that

$$plim\left[T^{-3/2}\sum_{t=1}^{T} u_t \left(\sum_{t=1}^{T} \varepsilon_s\right)\right] = 0 \tag{3.25}$$

It can be shown that $\psi T^{-3/2}\sum_{t=1}^{T} t u_t$ and $T^{-3/2}\sum_{t=1}^{T} x_t u_t$ are normally distributed with mean zero and variance $\psi^2\sigma_u^2/3$. Similarly, it can also be shown that $plim\left[T^{-3}\sum_{t=1}^{T} x_t^2\right] = \psi^2/3$. Hence, by Slutsky's theorem and Cramer's theorem,

$$T^{-3/2}(\hat{\beta} - \beta) = \left(T^{-3/2}\sum_{t=1}^{T} x_t u_t\right)\left(T^{-3}\sum_{t=1}^{T} x_t^2\right)^{-1} \xrightarrow{d} N(0, 3\sigma_u^2/\psi^2) \tag{3.26}$$

Therefore, when a deterministic trend is included, it dominates the stochastic trend asymptotically and the result is that the β coefficient will again have an asymptotically normal distribution.

Note that the asymptotic distribution caused by stochastic trends have particular implications for models relying on dummy variables, including dummy variables with longer runs that aim to capture structural change. The next section will explore this issue in greater detail, focusing as explained earlier, on price overcharge models.

3.4 Application: Price overcharge estimation

As set out in section 1.1, there are two main benchmark approaches used to estimate overcharge, namely the forecasting and dummy variable approaches. This dissertation mainly focuses on the dummy variable approach. I do, however, provide a brief discussion on the forecasting approach in the context of nonstationarity in Appendix D.

Similar to Chapter 2 consider a simple DGP of prices:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta x_t + \gamma p_{t-1} + \epsilon_t \tag{3.27}$$

where p_t is the product unit price in period t, D_t the cartel dummy variable, x_t is a set of explanatory variables and ϵ_t an error term. The sample period is $\mathbb{T} = \{1, ..., T\}$, which can be divided into periods with and without cartel effects, labelled $\mathbb{T}_C$ and $\mathbb{T}_N$ respectively, with $\mathbb{T} = \mathbb{T}_C \cup \mathbb{T}_N$.‘

The presence of unit roots introduces more nuance into the dummy variable approach. As previously mentioned, – in civil litigation – it is up to the plaintiff to prove the existence of overcharge. In contrast, it is in the interest of the defendant to argue that the overcharge was not significant. As a result, the statistical significance of the dummy variable coefficient in equation 3.27 is crucial to both parties. It was shown in section 3.3 that the intercept will no longer have an asymptotically normal distribution when dealing with unit roots. A dummy variable can be thought of as a shifting intercept. Therefore, the dummy variable coefficient will also not be asymptotically normally distributed. The results in section 3.6.1 illustrate the extent to which the distribution can differ from a Student's t-distribution and the related size distortions of the t-test. Hence, careful consideration should be given to the interpretation of the statistical significance of the dummy variable coefficient when any of the variables contain unit roots.

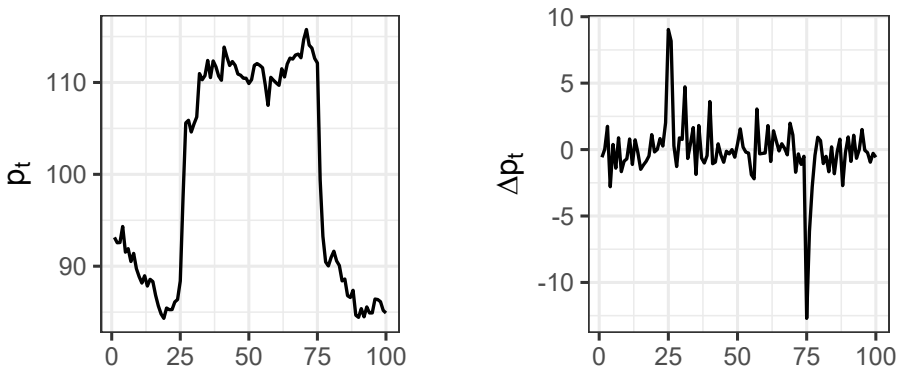
Aside from the issue of drawing inference on α_2 in equation 3.27, there are several estimation complications to be aware of. First, simply differencing the data to make the variables $I(0)$ is not an option. Differencing the data removes all long-run information and will completely remove the effects of the mean shift in p_t. The size of the mean shift in p_t is equivalent to the overcharge. Therefore, removing this effect makes estimating the overcharge impossible. Figure 3.1 illustrates graphically what would happen if the price series p_t was to be differenced to render it $I(0)$

Similar to the forecasting approach, when unit roots are present, the analysis needs to include cointegration tests to ensure that the relationships are not spurious. When dealing with single equations there are two types of cointegration tests available. The first type is the residual-based tests in the tradition of Engle and Granger (1987) and Phillips and Ouliaris (1990). The second type of test is the bounds-test of Pesaran *et al.* (2001). The residual-based test revolves around estimating a static cointegrating regression model of equation 3.27 which can be written as:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta x_t + v_t \tag{3.28}$$

and subsequently testing for unit roots in v_t. The bounds test would involve writing equation 3.27 as an autoregressive distributed lag model (ARDL) given by:

$$p_t = \alpha_0 + \omega D_t + \theta_0 x_t + \theta_1 x_{t-1} + \gamma p_{t-1} + u_t \tag{3.29}$$



(a) Simulated values of p_t (b) Difference of p_t

Figure 3.1: Example of the effects of differencing p_t

equation 3.29 can then be re-parameterised to present an error-correction model (ECM) as:

$$\Delta p_t = \theta_0 \Delta x_t - (1-\gamma)\left[p_{t-1} - \beta_0 - \beta_1 D_t - \beta_2 x_{t-1}\right] \tag{3.30}$$

$$\text{with } \beta_0 = \frac{\alpha_0}{1-\gamma}\ ,\ \beta_1 = \frac{\omega}{1-\gamma} \text{ and } \beta_2 = \frac{\theta_1}{1-\gamma}$$

The bounds-test involves estimating an unrestricted ARDL as[4]:

$$\Delta p_t = \alpha_0 + \psi D_t + \theta_0 \Delta x_t + \theta_0 x_{t-1} + \alpha p_{t-1} + u_t \tag{3.31}$$

Once equation 3.31 is estimated, an F-test is used to test if $H_0 : \theta_0 = \alpha = 0$. Note that x_{t-1} and p_{t-1} represent the long-run equilibrium, therefore, the test can be thought of as testing whether a long-run relationship exists.

There are two key issues to be concerned with regarding the dummy variable approach in the above described cointegration context. First, suppose that the Engle-Granger two step approach is followed. When estimating the static cointegrating regression model of equation 3.28, one cannot draw inference on α_2, since the regression contains $I(1)$ variables and the coefficient distributions will therefore be non-normal. In the Engle-Granger two-step approach the ECM can be written as:

$$p_t = \kappa_0 + \kappa_1 \Delta x_t - \pi \hat{v}_t + e_t \tag{3.32}$$

Now since $\hat{v}_t = p_t - \hat{\alpha_1} - \hat{\alpha_2} D_t - \hat{\beta} x_t \sim I(0)$, all the variables in equation 3.32 are $I(0)$ and standard inference applies. However, in the context of overcharge, this representation is not helpful since the overcharge coefficient is no longer in the model to draw inference from.

[4]Note that even though I use the same notation to represent the parameters as in equations 3.27 to 3.30, these parameters are not identical.

The second issue is how the dummy variable is handled when using an ARDL. Note that when using the ECM representation of equation 3.30, the estimation contains both $I(0)$ and $I(1)$ components and standard inference does not apply. An important problem arises in this case when estimating the unrestricted ARDL as in equation 3.31. When evaluated in this manner, the dummy variable is no longer restricted to only the long-run equilibrium. Therefore, the dummy variable is allowed to influence both the stationary and nonstationary part of the model. This implies that the structural break has both temporary and permanent effects on p_t. In terms of modelling overcharge, this is undesirable since it implies that during the collusion period the cartel raised prices and the price will never return to its competitive levels even after the disbandment of the cartel. It is therefore important that the dummy variable is restricted to only the long-run relation.

The following questions are therefore important to address when unit roots are present in the data. When using the dummy variable approach, inference regarding the significance of the dummy variable coefficient using Student's t-statistics cannot be made. To illustrate why caution needs to be applied, I empirically determine the difference between the true distribution of the dummy variable coefficient and a Student's t-distribution. I also illustrate the size distortions that occur when using t-statistics.

The implication of the above is that the dummy variable should be restricted to the long-run equilibrium relationship and that standard inference is biased. There is a further problem, related to the insights of the previous chapter. Up to this point, I have considered the break dates to be known. The break dates inform the construction of the dummy variable as discussed in Chapter 1 and 2. In Chapter 2 I illustrated the importance of correctly determining the break dates. Therefore, a relevant question is, How cointegration tests perform when the break dates are misspecified? Put differently, to what extent is the solution of the unit root problem dependent on the appropriate dating of the structural break? Cointegration tests form the basis of concluding whether a long-run relationship exists or whether the results are spurious. As discussed in Chapter 2 the legally defined infringement period does not always coincide with the effective period. It is therefore important to know how sensitive cointegration tests are to the specification of the break dates. I show below that incorrect conclusions, with regard to the existence of the long-run relation, are likely when the break dates are misspecified. In particular I test the performance of residual-based cointegration tests and the bounds-test when the break dates are misspecified.

3.5 Methodology

By performing Monte Carlo simulations in conjunction with time series bootstraps I can obtain the distribution of the dummy variable coefficient for var-

ious sample sizes. As illustrated in section 3.3, the distribution of the dummy variable coefficient in a cointegrated regression will converge asymptotically to a stochastic variable. Therefore, the distributional form of the dummy variable coefficient will be random in finite samples. As such, empirically illustrating the distributions in various simulations will not provide a useful result for practitioners. I therefore extend the results by illustrating the magnitude of the difference between the "true" distribution and Student's t-distribution. By presenting the simulation results in this manner, it is easier to understand when incorrect inferential conclusions can be made when not considering the effects of unit roots.

Given that the true distribution of the dummy variable coefficient differs from that of a Student's t-distribution, it is important to quantify the probability of making type I and II errors. To explore this issue, I make use of p-value plots. These graphs plot the number of p-values smaller than a certain threshold against the threshold where the p-value is constructed, based on the Student's t-distribution. Presented in this way, the plots allow assessment of whether a t-test will systematically over- or under-reject the null hypotheses.

The methodology can be described as follows. I generate a cointegrated system that contains a dummy variable and perform time series bootstrapping to obtain the true distribution of the dummy variable coefficient. I then calculate the statistical difference between the true distribution of the coefficient and that of a t-distribution using the Kolmogorov-Smirnov test statistic. This process is repeated 10 000 times to ensure that the difference between the distributions is systematic. It is well known that increasing the number of repetitions in a MC experiment will increase the accuracy of the results. However, given the computation intensity of the methodology in this chapter, the number of repetitions has to be limited. Following the convention of Engle and Granger (1987), Engle and Yoo (1987), Schwert (2002) and Phillips and Ouliaris (1990) I limit this number to 10 000. Furthermore, I deem this number of replications sufficient given that preliminary results showed that there is no difference in the conclusions drawn when 5 000 repetitions is used. The simulation is repeated for different sample sizes T and different sizes of the dummy variable D_t.

This section is divided into four subsections. In section 3.5.1 the data simulation and use of the Monte Carlo method is explained. As outlined in section 3.3, the problem of nonstationary data is that the coefficient of the dummy variable will no longer have a Student's t-distribution. The results, therefore, show the differences between a Student's t-distribution and the true distribution of the dummy variable when nonstationary data is used. The true distribution of the dummy variable is obtained through a maximum entropy (ME) bootstrap for which the methodology is explained in section 3.5.2. The process followed to calculate the differences between the bootstrapped distribution and the Student's t-distribution is explained in section 3.5.3. In section 3.5.6 I explain the methodology followed to display how the results translate

to the practical application of overcharge.

Given the relative complexity of the simulation study, Table 3.1 contains the step-by-step procedure that was followed.

Table 3.1: Pseudo code for simulation study

1. Generate a single DGP with fixed length of T and D_t.
2. Obtain the bootstrapped distribution of the dummy variable coefficient (α_2) using ME bootstrap.
3. Compare the distribution of (2) with a t-distribution using the K-S test statistic.
4. Repeat steps (1) to (3) 10 000 times.
5. Change only the length of D_t while keeping the size of T the same. Repeat steps (1) to (4) while incrementally increasing the size of D_t.
6. Incrementally increase the size of T while repeating steps (1) to (5).

3.5.1 Data simulation

The data is generated through the following data generating processes:

$$y_t = \alpha_1 + \alpha_2 D_t + \beta x_t + \gamma y_{t-1} + u_t \tag{3.33}$$

$$x_t = x_{t-1} + \varepsilon_t \tag{3.34}$$

$$u_t \sim IN(0, \sigma_u^2), \quad \varepsilon_t \sim IN(0, \sigma_\varepsilon^2)$$

In each repeated sample, the coefficient values of equation 3.33 is fixed. x_t is a random walk and as a result y_t will be nonstationary. Since $u_t \sim IN(0, \sigma_u^2)$, it follows that the imposed linear combination of y_t and x_t is cointegrated. This form of specification reflects the type of reduced-form regression model employed in overcharge estimations.

I set $\alpha_1 = 100(1 - \gamma)$, so that the mean price level in the simulations is 100. I furthermore choose $\beta = 0.6$ and $\alpha_2 = 10$. I find that the results are not sensitive to the choice of the coefficient estimates. In section 3.5.6 I will discuss a more precise match with the cartel damages literature.

Monte Carlo simulations can be used to computationally mimic the sampling distributions of estimators and test statistics[5]. Computing the test statistics with Monte Carlo simulations is critical to the exercise. From equation

[5]For more in-depth and technical discussions of Monte Carlo methods see Gentle (2003) and Dunn and Shultis (2012)

3.21 in section 3.3, the asymptotic distribution depends heavily on the relationship between x_t and u_t. It is therefore important to repeat the simulation – while keeping the coefficient values and sample size (T) constant – in order to obtain a large range of relationships between x_t and u_t for each of the regressions. This ensures that a wide range of plausible values of x_t is considered and that the results are not dependent on the values of x_t. Furthermore, the difference in distribution between the bootstrap and Student's t-distribution is calculated using the sample data. This yields an estimate that could be biased or inefficient. Therefore, it is important to compute these statistics in a Monte Carlo setting to ensure no systematic bias or inefficiency is present.

The above simulation is repeated 10 000 times for fixed coefficient values, sample size and dummy variable length. Subsequently, the entire simulation is also repeated by incrementally increasing the length of the dummy variable while keeping the coefficients and sample size fixed. I then revert to a shorter dummy variable length and increase the sample size. The procedure, therefore, involves simulating various regressions for fixed coefficient values while varying the length of the dummy variable and sample size.

In the following section, I elaborate on how the time series bootstrap is obtained for each regression.

3.5.2 Time series bootstrap

For the simulation study, I require a method to empirically derive the distribution of the dummy variable coefficient for each of the simulated regressions. To obtain the empirical distribution for a single sample, I make use of a time series bootstrap. There are two general methods that may be used to bootstrap a regression. The two methods involve treating the predictors either as fixed (*case resampling*) or random (model-based resampling). To provide a simple explanation of the two methods, consider the DGP from section 3.5.1:

$$y_t = \alpha_1 + \alpha_2 D_t + \beta x_t + \gamma y_{t-1} + \epsilon_t \tag{3.35}$$

Case resampling involves simultaneously resampling (y_t, x_t). For each sample of (y_t, x_t), I fit the regression from equation 3.35 and save the coefficient values. By repeating this procedure a large number of times, the empirical distributions of the coefficients can be obtained from the saved coefficient values. With model-based resampling, I would first run the regression on the observed data and obtain the fitted residuals $\hat{\epsilon}_t$. In the following step, the values of the predictor(s) (x_t) are assumed to be fixed and resamples are created for $\hat{\epsilon}_t$. For each of the resampled residual series $(\hat{\epsilon}_t)$ I refit the regression from equation 3.35 and save the coefficient values for each refitted regression. The empirical coefficient distributions are then obtained in the same way as with case resampling.

A simple conceptualization of the difference between the two above-mentioned methods is that case resampling treats the predictors as random while model-

based resampling treats the predictors as fixed and the dependant variable as random. The model-based resampling procedure relies heavily on the model specification. Model misspecification would result in an empirical distribution that would be different from the true distribution of the coefficients. It is also important to note that model-based resampling requires explicit assumptions to be made regarding the distributional form of the residual series.

In this chapter, I used case resampling to obtain the results. Aside from longer computing times required, compared to model-based resampling, there are a few advantages to following this approach (Davidson and Hinkley, 1997). First, this method is not influenced by the assumed reliability of the specified parametric model. Second, it does not assume that the conditional mean of the dependant variable given the realizations of the predictors is linear. Third, it still provides robust results with heteroskedastic errors. The first advantage builds a strong case for why case resampling should be favoured in applied work where there can be uncertainty regarding model specification. I favoured case resampling since there is a lack of literature on the properties of bootstrapping on cointegrated series with dummy variables. I therefore preferred an approach that does not require us to make *a priori* assumptions regarding the distributional form of the residual series and that can accommodate heteroskedastic residuals that may be present in small samples. Furthermore, an issue regarding the treatment of structural changes – that is relevant to this application – has gone unnoticed by the literature and is addressed in this section.

The bootstrap is a widely used non-parametric method for approximating the sampling distribution of complicated statistics based on independent identically distributed (iid) observations. Time series data – especially when the data is nonstationary – is typically not independent identically distributed. Since this methodology is centred around nonstationary data, the iid assumption of bootstrapping will be violated.

To overcome the violation of the iid assumption, a maximum entropy bootstrap is used, as outlined in Vinod (2004, 2006). The methodology is as follows.

Let $f(x)$ denote the density function of a time series x_t with $t = 1, ..., T$. The entropy H is defined as:

$$H = E(-logf(x)) \tag{3.36}$$

The first step is to re-order the data x_t in increasing order and saving the ordering index. The order statistics of x_t is denoted as $x_{(t)}$. After re-ordering, intermediate points z_t are obtained from the order statistics where

$$z_t = 0.5 * (x_{(t)} + x_{(t+1)}) \tag{3.37}$$

Next, calculate the trimmed mean m_{trm} of absolute distances between consecutive observations, $|x_t - x_{t-1}|$. The lower limit (L) for the left tail and upper limit (U) for the right tail is obtained from

$$z_L = x_{(1)} - m_{trm} \tag{3.38}$$
$$z_U = x_{(T)} + m_{trm} \tag{3.39}$$

The limits of z_L and z_U become the limiting intermediate points of z_t from equation 3.37. Using the midpoints z_t define half open intervals as:

$$I_t = (z_{t-1}, z_t], \ t = 1, ..., T \tag{3.40}$$

This will result in T intervals of I_t each of which contains one element of $x_{(t)}$. I_t will be used to form a re-sample. The re-sample will contain one observation from each interval with probability $1/T$.

For each interval I_t the mean of the maximum entropy (ME) density is calculated such that the 'mean-preserving constraint'[6] is satisfied. The ME density is defined as the combination of T uniform densities over T half open intervals.

After calculating the ME density, random numbers from the $[0; 1]$ uniform intervals are independently drawn to compute quantiles of the ME density. The quantiles are then re-ordered by using the ordering index that was obtained from the first step. This will reinstate the time-dependence relationship that is in the original observed data (x_t).

Following the steps outlined above will provide a re-sampled series, (x_t^*) of x_t that relates to the corresponding values of y_t. To obtain the approximate distribution of the dummy variable coefficient, the following procedure is employed. First, consider the regression that is run on the original data x_t:

$$y_t = \alpha_1 + \alpha_2 D_t + \beta x_t + \gamma y_{t-1} + \epsilon_t \tag{3.41}$$

To obtain the ME bootstrap for the coefficient α_2 I re-estimate equation 3.41 using the re-sampled series x_t^*. The value of α_2 obtained from using the re-sampled data is then stored. The entire process is then repeated 10 000 times with each run re-sampling x_t, estimating the regression equation 3.41 and storing the value of α_2. The final result is a vector of 10 000 α_2's. From this vector, the distribution of α_2 is constructed.

Specific to this case, special care needs to be taken when performing a *case resampling* bootstrap in a model with dummy variables. The following discussion attempts to make a modest but important contribution to the bootstrap

[6]The constraint states that the means for the uniform density function, m_t, must satisfy the following relations:

$$\begin{aligned} m_1 &= 0.75x_{(1)} + 0.25x_{(2)}, \\ m_k &= 0.25x_{(k-1)} + 0.5x_{(k)} + 0.25x_{(k+1)}, \text{ for } k = 2, ..., T-1 \\ m_T &= 0.25x_{(T-1)} + 0.75x_{(T)} \end{aligned}$$

literature. When dealing with level shifts or structural breaks that require the use of a dummy variable, it is important to be mindful of the position of the break in the series when re-sampling is done during the bootstrapping procedure. Re-sampling will slightly shift the positions of x_t and y_t values around the break or shift. If values of the break or shift are moved to an outside regime the subsequent regression on the re-sampled series will have biased estimates as a result of uncontrolled outliers. Figure 3.2 graphically illustrates this problem. The grey area represents the true regime in which the originally specified regression contained an intercept shift.

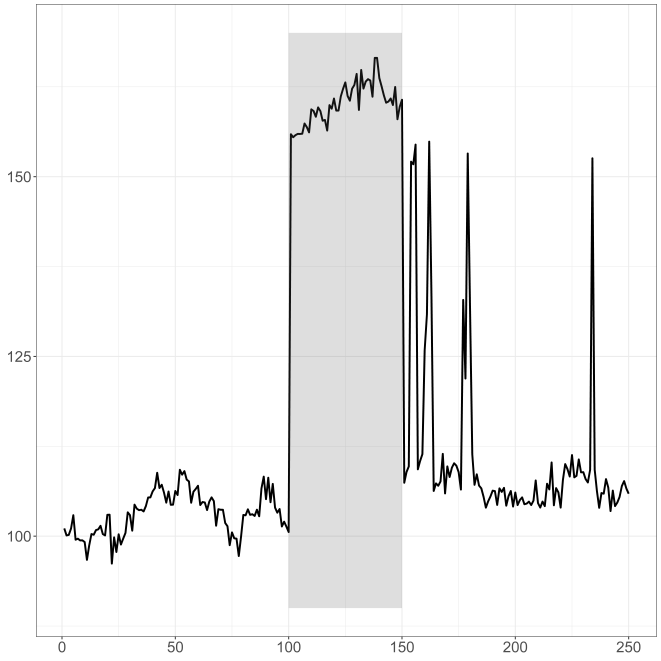


Figure 3.2: Unrestricted bootstrap with breaks

From the graphic, it is clear that some periods contain outliers that are not controlled for when running the subsequent regressions. It is therefore important to restrict the re-sampling to only re-sample from periods that belong to the same regime. Note that this problem is not specific to the ME bootstrap illustrated here. Using the block bootstrap from Politis and Romano (1994) or similar methods will have the same problem, and care should be taken to restrict the re-sampling as well.

3.5.3 Obtaining differences between distributions

Contrary to previous work, I are not as interested in approximating the asymptotic distributions for various coefficients. Instead, I choose to focus on how the coefficient distribution differs from that of a Student's t-distribution in

finite samples. For applied work, it is essential to know the magnitude of this difference between distributions, since it illustrates the likelihood of making type I or II errors. To test the magnitude of difference between the bootstrapped distribution and the t-distribution, the Kolmogorov–Smirnov (K-S) (Kolmogorov, 1933; Smirnov, 1948) test is used. The K-S statistic quantifies the distance between the bootstrapped distribution and the t-distribution.

The K-S test statistic is obtained as follows. Let X_i be n independent and identically distributed ordered observations. The empirical distribution function $F_n(x)$ for X_i is then defined as:

$$F_n(x) = 1/n \sum_{i=1}^{n} I_{[-\infty,x]}(X_i) \tag{3.42}$$

where $I_{[-\infty,x]}(X_i)$ is an indicator function that is equal to 1 if $X_i \leq x$ and equal to 0 otherwise. For a given cumulative distribution function $F(x)$, the K-S statistic is given by:

$$D_n = \underset{x}{sup}|F_n(x) - F(x)| \tag{3.43}$$

where $\underset{x}{sup}$ is the supremum of the set of distances. The H_0 is that the empirical distribution $F_n(x)$, the bootstrapped distribution, is the same as the specified distribution $F(x)$, the t-distribution. If the distributions are similar, the test statistic D_n will converge to 0. I report both the test statistics and number of rejections for the Monte Carlo simulations.

3.5.4 Size distortions

The K-S results illustrate how the empirical distribution of the dummy variable differs from the theoretical distribution. More formally, the K-S tests indicate whether there is a statistically significant difference between the empirical and theoretical distributions. However, for practical applications, knowing the magnitude of the difference between the true and theoretical distributions is not particularly useful. What is of greater interest is the size of the t-test on the dummy variable coefficient. The size of the test statistic is the probability of making a type I error, that is the probability of rejecting $H_0 : \alpha_2 = 0$ when H_0 is actually true. To illustrate the size distortions that occur when testing the significance of the dummy variable, I use p-value plots.

P-value plots provide a simple method to distinguish between test statistics that systematically under-reject or over-reject H_0. Since the plots convey information about the probability of a test over-rejecting or under-rejecting, it is an intuitive graphical method to evaluate the size of the test at various levels of significance.

To obtain the p-value plots, I perform a Monte Carlo experiment in which N realizations of the test statistic τ are generated where each value of the test

statistic is denoted as τ_i. The test statistic, τ_i, is the scaled t-statistic of the dummy variable coefficient. The p-value of τ_i is the probability of observing a value of τ more extreme or as extreme as τ_i assuming that $H_0 : \alpha_2 = 0$ is true according to some distribution $F(\tau)$. In this case, $F(\tau)$ is the corresponding t-distribution of τ. Using the scaled t-statistic and corresponding t-distribution I can find the p-value of τ_i which is denoted as $p_j \equiv p(\tau_i)$. To construct the p-value plots, I require the empirical distribution function (EDF) of p_j which is an estimate of the cumulative distribution function of $p(\tau)$. The EDF of p_j at any point on the interval $x_j \in (0, 1)$ is given by

$$\hat{F}(x_j) = \frac{1}{N} \sum_{i=1}^{N} I(p_j \leq x_j) \tag{3.44}$$

where $I(p_j \leq x_j)$ is an indicator function that is equal to 1 if the inequality is true, and 0 otherwise.

The graph plots $\hat{F}(x_j)$ against x_j. By constructing $\hat{F}(x_j)$ using equation 3.44 it follows that if the distribution of τ that was used to compute each p_j was correct, then each of the p_j should be distributed as uniform $(0, 1)$. Therefore, when $\hat{F}(x_j)$ is plotted against x_j the result should be close to the 45^0 line.

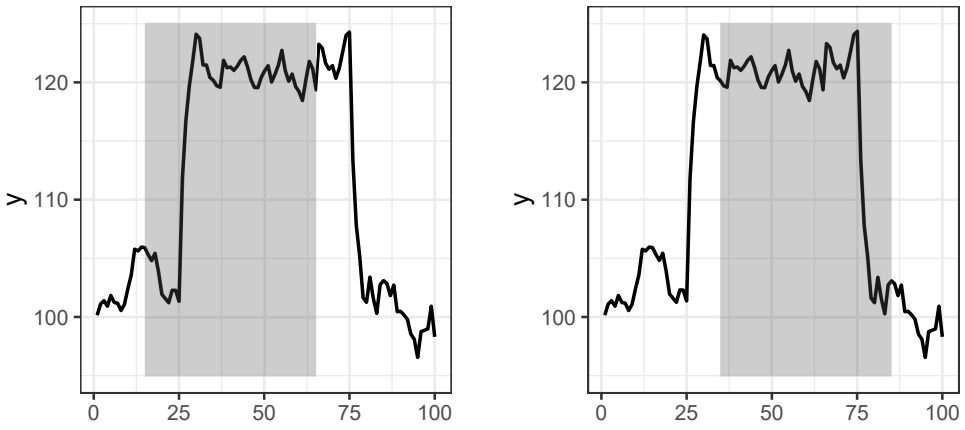
Due to the super-consistency of $\hat{\alpha_2}$, there will be little variance between bootstrapped values of $\hat{\alpha_2}$ as a result of the variance in x_t. Instead, the $var(\hat{\alpha_2})$ between bootstrapped samples will depend on the error variance, $var(\epsilon_t) = \omega$. For larger values of ω, the variance between bootstrapped values of $\hat{\alpha_2}$ will be larger. If ω is small, the variance of $\hat{\alpha_2}$ between bootstrapped samples will be small, and the t-statistic will also be small in each bootstrapped sample. That is, the bootstrapped values of t will all be tightly centred around 0 when ω is small.

As explained in section 3.3, the distribution of the dummy variable coefficient (α_2), depends on the relationship of the error terms of x_t and y_t, denoted as ζ. In equation 3.22 I showed that when $\gamma = 0$ the coefficients will be normally distributed. Therefore, the size distortion is dependent on the value of ζ. I construct the p-value plots for different values of ζ to illustrate the respective size distortions of the t-test when the error terms of x_t and y_t have a strong and weak relationship.

3.5.5 Cointegration tests

As discussed earlier, cointegration modelling does not necessarily solve the problem posed by stochastic trends in finite samples. In particular, when a structural break is incorrectly dated, new problems may arise. To illustrate the complications of cointegration tests in the presence of mean shifts, I simulate data with the DGP given as in equation 3.33. For each of these simulations, I fit two types of regression models. First, I fit a model where the misspecified

dummy is placed at the start of the sample, and second, I fit a model where the misspecified dummy lies towards the end of the sample. Figure 3.3 displays an example of this type of misspecification, with the grey area displaying the observations for which the dummy variable would equal 1.



(a) Misspecification at the start of the sample

(b) Misspecification at the end of the sample

Figure 3.3: Misspecification dummy encoding example

For these simulations, I consider an intermediate sample of $T = 100$. In the simulation, I have that $D_t = 1$ *for* $t = 25, ..., 75$ and $D_t = 0$ elsewhere. I define T_m as the misspecification percentage. I consider three different values of T_m, being 10%,20% and 50%. For example, if $T_m = 20\%$, and I consider misspecification at the start of the sample, I would incorrectly encode the dummy variable in the regression as $D_{t_m} = 1$ *for* $t = 15, ..., 65$ and 0 elsewhere, as is illustrated in Figure 3.3 (a).

Given that I have a single equation, there are two types of cointegration tests that require consideration. The first is the residual-based test as set out in Granger (1981). This is a simple case of testing the residuals for a unit root where the test statistics are given in MacKinnon (1991). The second cointegration test is the bounds test of Pesaran *et al.* (2001).

3.5.6 Application to overcharge estimation

The focus of this discussion is on cartel overcharge estimation. The simulation – as outlined in section 3.5.1,3.5.2 and 3.5.3 – should therefore be adjusted.

The functional form of the data simulation is changed to:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta_1 c_t + \beta_2 d_t + \gamma p_{t-1} + \epsilon_t \tag{3.45}$$
$$c_t = c_{t-1} + v_t \tag{3.46}$$
$$d_t = d_{t-1} + a_t \tag{3.47}$$
$$\epsilon_t \sim IN(0, \sigma_\varepsilon^2), \quad v_t \sim IN(0, \sigma_\varepsilon^2), \quad a_t \sim IN(0, \sigma_\varepsilon^2)$$

where c_t and d_t represent cost and demand drivers of price, p_t, and D_t is representative of the collusive period. Importantly, c_t and d_t are restricted not to contain any negative values. This does not mean that c_t and d_t cannot have downward trends. Instead, it implies only that the value of c_t and d_t cannot be below zero at any given t. The restriction is placed since the cost and demand drivers of price cannot be negative in practice.

As explained in section 3.4, a primary concern in applied work is the specification and significance of the dummy variable D_t. To illustrate this problem, I deliberately misspecify the dummy variable where 50% of its values do not correspond to the true break dates. For example, if the intercept in the DGP occurred at $t = 20$ to $t = 30$ the dummy is encoded to contain 1s from $t = 15$ to $t = 25$ and 0s everywhere else. I then test the number of times in each of the simulations that the corresponding t-statistic would have resulted in concluding that the misspecified dummy variable's coefficient is significant.

3.6 Results

This section consists of two subsections. In section 3.6.1 the results show how much the true distribution of the dummy variable coefficient differs from a Student's t-distribution when nonstationary data is used. In section 3.6.2 I investigate the size distortions of using t-tests when the data contains unit roots. Section 3.6.3 evaluates the performance of cointegration tests when the break dates are misdated. Section 3.6.4 illustrates how the above-mentioned problem translates to an applied study of overcharge.

3.6.1 Effects on inference

To illustrate the effects on inference I perform Monte Carlo simulations as outlined in section 3.5.1. The simulation contains a dummy variable and the results compare the empirical distribution to that of a t-distribution. That is, for each of the series, I obtain the bootstrapped distribution and compare it with the t-distribution using the K-S statistic. The simulation is then repeated for various sample sizes T and dummy variable lengths. Figure 3.4 shows the bootstrapped distribution of the dummy variable plotted over the Student's t-distribution for $T = 100$ and $T_c = 20$. That is, the 'true' distribution of the dummy variable coefficient against that of a t-distribution for a single simulated series.

The difference between the left and right tail illustrates the problem faced when not appropriately considering the effects of nonstationary data. Since the Student's t-distribution has a fatter left and right tail, i.e. the distribution is leptokurtic, there is a higher probability that an analyst is likely to conclude that the dummy variable coefficient is significant when it is not. Making type I errors is more likely due to the larger rejection region of the t-distribution compared with the bootstrapped distribution.

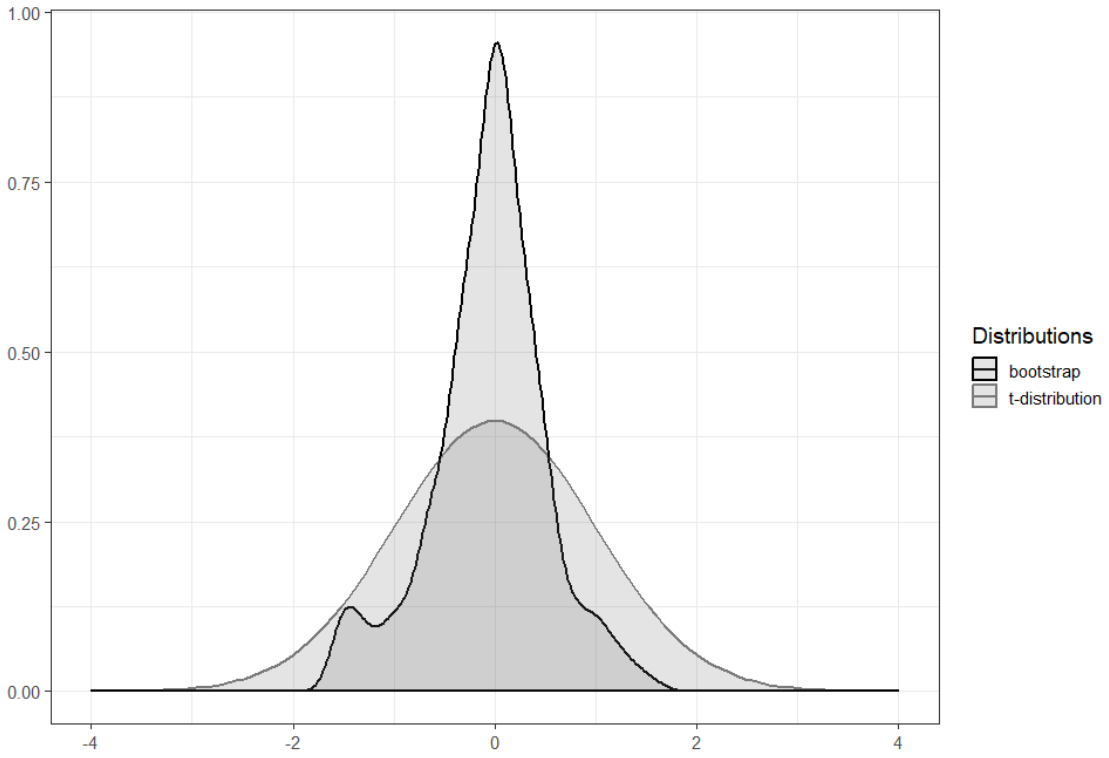


Figure 3.4: Difference between bootstrapped-distribution and t-distribution

The distribution of the dummy variable depends heavily on the sample size (T) and the length of the dummy variable (T_c). The result of interest, therefore, is how much the bootstrapped distribution differs from that of a t-distribution. Hence, I calculated the difference between the dummy variable bootstrapped distribution and t-distribution for various values of T and T_c using the K-S statistic. Figure 3.5 represents the median value of the K-S statistic.

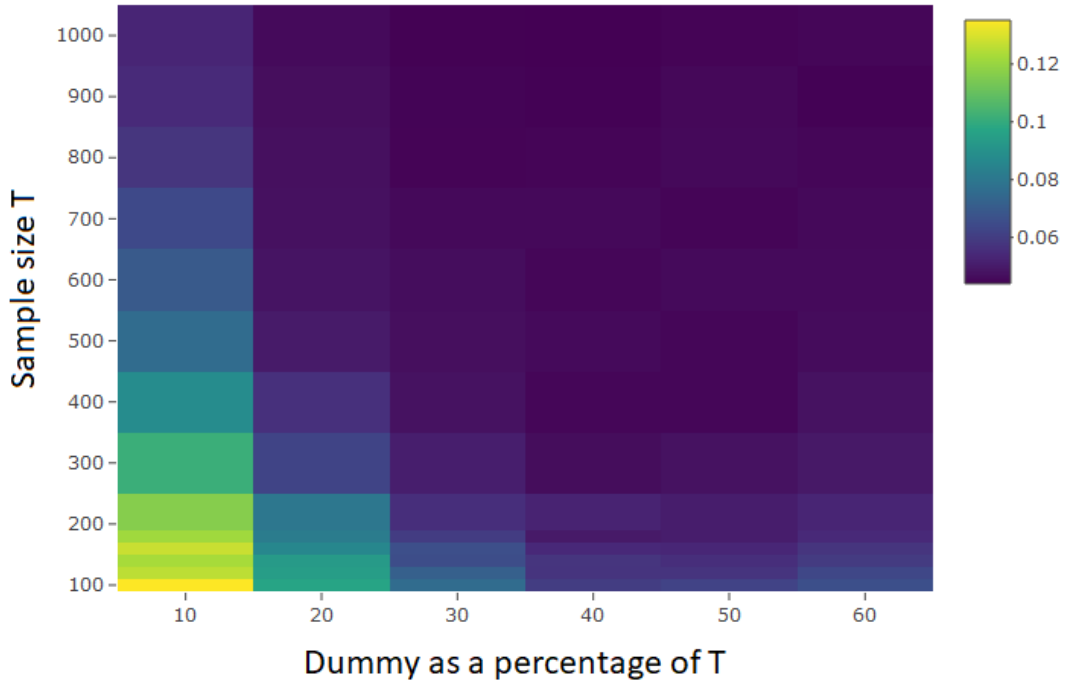


Figure 3.5: Median K-S statistic between bootstrapped-distribution and t-distribution

The dummy variable length (x-axis) is reported as a percentage of the total sample size T. For example, if the dummy variable length is reported as 10% when the total sample size is $T = 100$ the implication is that 10 periods out of 100 contain a 1 for the dummy variable. From section 3.5.3, the value of the K-S statistic would be close to 0 if the two distributions are similar. Therefore, larger values imply that the magnitude of difference between the two distributions is larger. Note that each combination of T and dummy variable as a percentage of T contains 10 000 simulations where the bootstrapped distribution and K-S statistic is calculated for each one of the 10 000 series.

The results in Figure 3.5 are useful for applied work involving data with stochastic trends. It shows that the difference between the "true" distribution of the dummy variable coefficient and a t-distribution is at its largest when the sample size is small and the dummy variable is small in relation to the sample size. Therefore, if the sample size is small and the dummy variable makes up a small portion of the total sample, there is an increased likelihood of drawing incorrect inference when erroneously using a t-distribution.

The median value may be inaccurate if the vast majority of the K-S statistic values are far away from the centred value. Therefore, Figure 3.6 shows the

percentage of the 10 000 simulations for each combination of T and T_c, for which the K-S statistic was statistically significant.

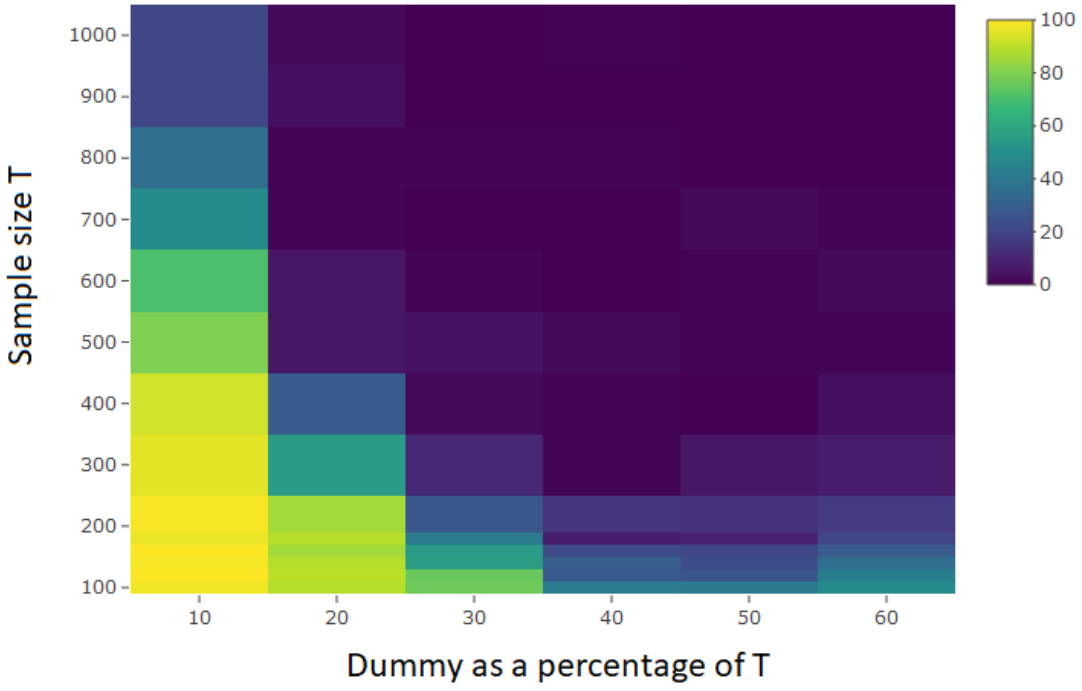


Figure 3.6: Significance of K-S test statistic

Similar to the result in Figure 3.5, the difference between the distributions is significant for almost 100% of the 10 000 simulations for each combination of T and T_c. This difference remains significant for a high percentage of the simulations for large samples when T_c is small. Although, when evaluating Figures 3.5 and 3.6 together, the result is less concerning for large samples, since the median magnitude of difference between the distributions seems to be relatively small.

When dealing with smaller samples and nonstationary data, careful consideration should be given to the distribution of the dummy variable coefficient. The shape of the distribution is much more nuanced than those associated with the coefficients of other regressors. Furthermore, the differences between distributions will be more severe when using real economic data. This is because in the simulation the relationship between the residual and dummy variable is extremely small due to the control that I have over the DGP. In applied work, it is more likely that the assumption of strict exogeneity needs to be relaxed, in which case the difference in distributions will be much larger than what is reported here.

3.6.2 Test size

In this section, I make use of p-value plots to illustrate the size distortion that occurs when using t-tests. The figures plot $\hat{F}(x_j)$ against x_j, where $\hat{F}(x_j)$ is the EDF of the p-values obtained from the t-statistic of the dummy variable coefficient, and x_j is the level of significance. Note that if the distribution used to construct the p-values was correct then each of the p-values should be distributed as uniform $(0, 1)$. Therefore, when $\hat{F}(x_j)$ is plotted against x_j, I expect the values to be close to the 45^0 line. In other words, I expect the number of occurrences where the p-value is smaller than the significance level x_j to be equal to the significance level x_j.

As explained in section 3.3, the asymptotic relation between the normal distribution and the dummy variable coefficient depends on ζ, where ζ is the relationship between the error terms of x_t and y_t. To compute the size distortions of the t-tests, I require this relationship to be fixed. This is because allowing it to vary will cause the shape of the EDF to be different between MC simulation runs.

I compute $\hat{F}(x_j)$ for three different values of ζ: $\zeta = 0.01$, $\zeta = 0.5$, $\zeta = 0.8$. For each value of ζ I vary the sample size T and the relative size of the dummy variable coefficient T_c. The results are reported in Figures 3.7 to 3.10.

As expected, when $\zeta > 0.01$, the t-test tend to over-reject the null hypothesis since the values are far above the 45^0 line. The sample size T plays a small role in whether the test over-rejects the null hypothesis. For example, the curves for different values of ζ in Figure 3.7 when $T = 100$ are marginally higher than the curves in Figure 3.8 when $T = 100$. However, consistent with the results in section 3.6.1, the relative size of the dummy variable T_c plays a significant role in whether the test over-rejects. For example, in Figure 3.7 the curves are much closer to the 45^0 line when $T_c = 0.5T$ than when $T_c = 0.1T$. This is because, as was shown in section 3.6.1, the difference between the t-distribution and true distribution of the dummy variable coefficient is much smaller when T_c spans over a longer period of T.

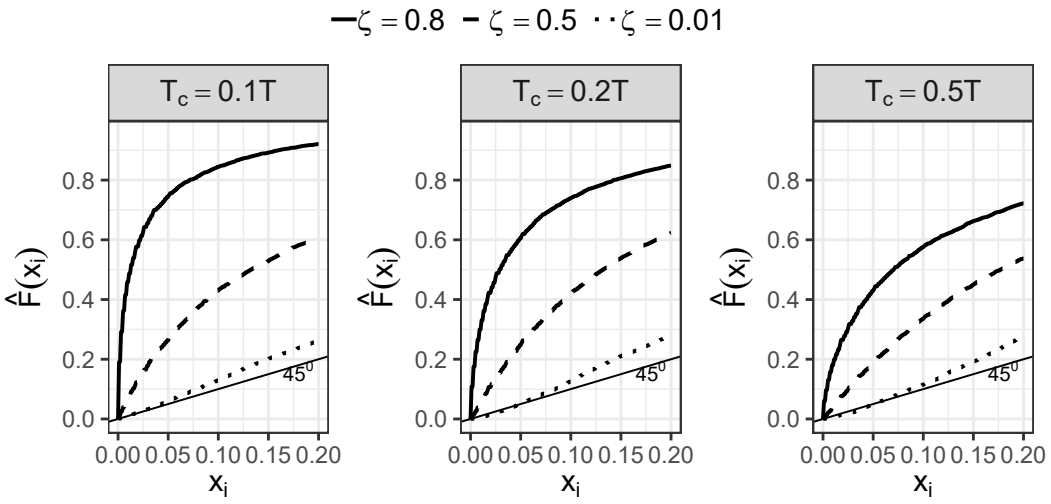


Figure 3.7: P-value plot, $T = 50$

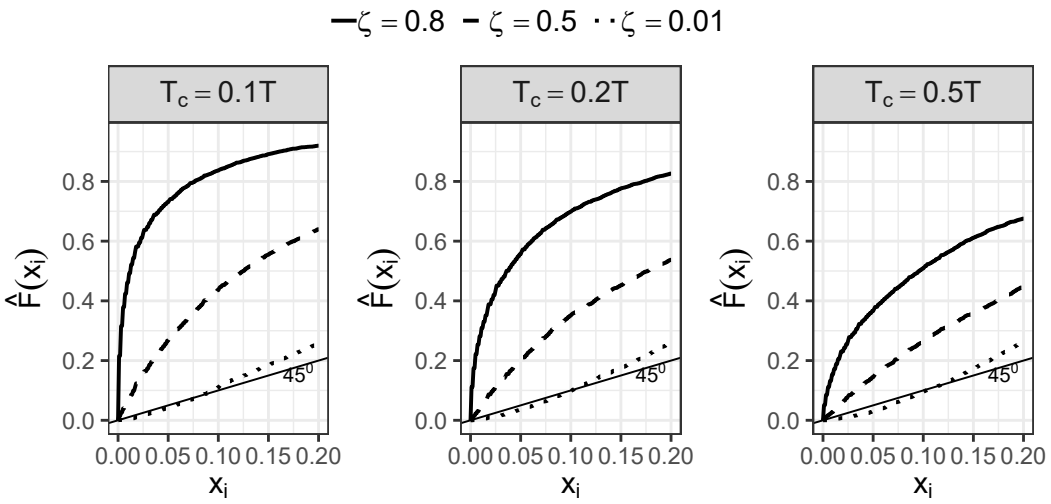


Figure 3.8: P-value plot, $T = 100$

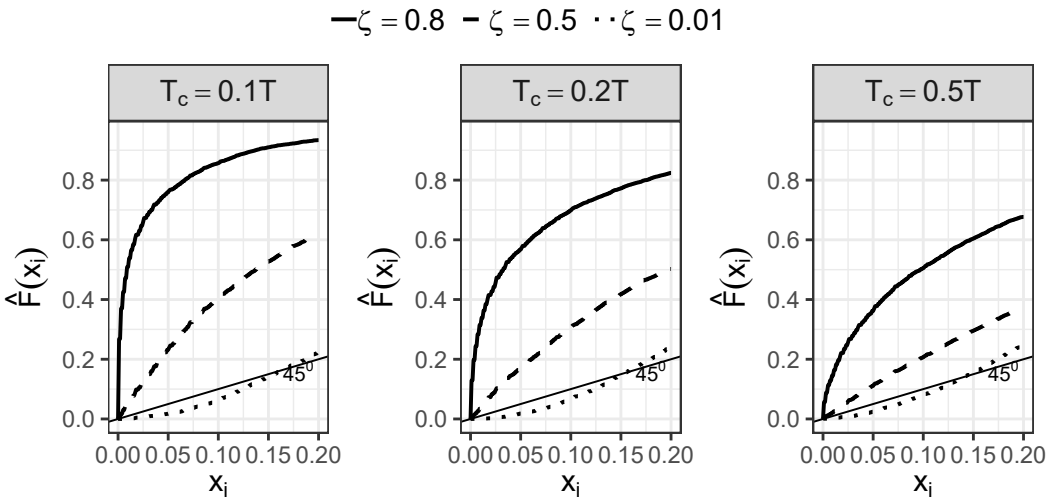


Figure 3.9: P-value plot, $T = 200$

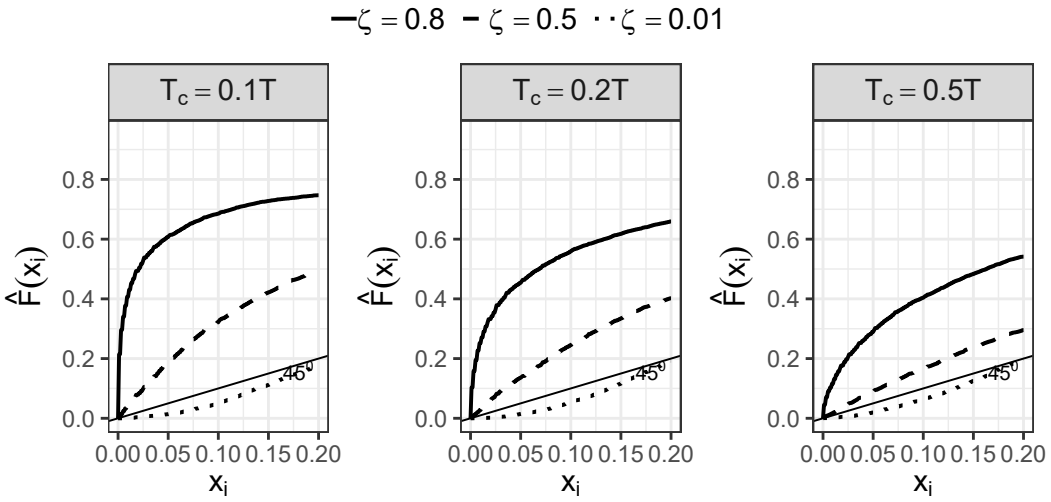


Figure 3.10: P-value plot, $T = 1000$

3.6.3 Effect on cointegration testing

Cointegration modelling is often considered as a solution to some of the problems associated with unit roots. It could therefore be considered a solution when dealing with dummy variables. However, as argued earlier in this chapter, other problems emerge, especially when structural break dates are not accurately captured by the dummy variable. Note that I focus on a specific case where there is a single cartel period. Therefore, there are two break dates

that correspond with the start and end date of the cartel. The dummy variable therefore contains a single set of repeated 1s and has 0s elsewhere.

As formally set out in section 3.5.5, I investigate the performance of the residual- based and bounds cointegration tests. I test the performance of both tests when misdating occurs at the start and end of the true break period[7]. When misdating is at the start of the true break period the 1's in the dummy variable overlap with the period before the first break date. Conversely, when misdating is at the end of the true break period the 1s in the dummy variable overlap with the period after the second break date. Additionally, I investigate the effect of various sizes of misdating, where misdating is defined as the proportion of 1s that incorrectly overlap with the pre- and post- break periods.

Note that for the residual-based test, the null hypothesis is that no cointegrating relationship exists. The null is rejected when the test statistic is smaller than the critical value. For the bounds test, Pesaran *et al.* (2001) provide lower and upper bounds on the critical values. If the F-statistic is larger than the upper bound, the null hypothesis is rejected and the conclusion is that there is cointegration. If the F-statistic is smaller than the lower bound the null hypothesis cannot be rejected, and the conclusion is that there is no cointegration. Lastly, if the F-statistic falls between the upper and lower bound the test is inconclusive.

The residual-based test has a 100% rejection rate of the null of no cointegration. This occurs for all three levels of significance ($\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$) and does not depend on whether the misdating is at the start or end of the sample. The results, therefore, are not reported in tabulated format. I provide a discussion on why the residual-based test has a 100% rejection rate in Appendix E.

Table 3.2 and 3.3 show the proportion of rejections of the null for the bounds tests when misspecification is at the beginning and end of the sample respectively. Furthermore, table 3.2 and 3.3 include a benchmark case of no misspecification. In this case the dummy variable is well specified in relation to the break periods. The results in Tables 3.2 and 3.3 can be interpreted as the power of the test. As mentioned above, rejecting the null implies that there is a cointegrating relationship. By construction y_t and x_t is cointegrated, therefore the tables report the percentage of occurrences where the test rejected the null when the null is false.

From Table 3.2, it can be seen that the bounds tests has relatively good power when the misdating occurs at the start of the true break period and the percentage of misspecification is small. As expected, when the size of misdating increases the power of the bounds test falls significantly. This is because the bounds test is dependant on the estimated long-run coefficients. As the size of the misdating increases, the bias in the long-run coefficients will increase.

[7]For a graphical explanation of misdating at the start or end of the true break period see figure 3.3.

Table 3.2: Percentage of rejection of H_0 when misspecification is at the start of the sample

	Significance level		
Misspecification percentage	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
10%	63%	92.88%	98.02%
20%	7.5%	26.5%	44.3%
50%	4.44%	10.88%	16.3%
no misspecification	99.99%	100%	100%

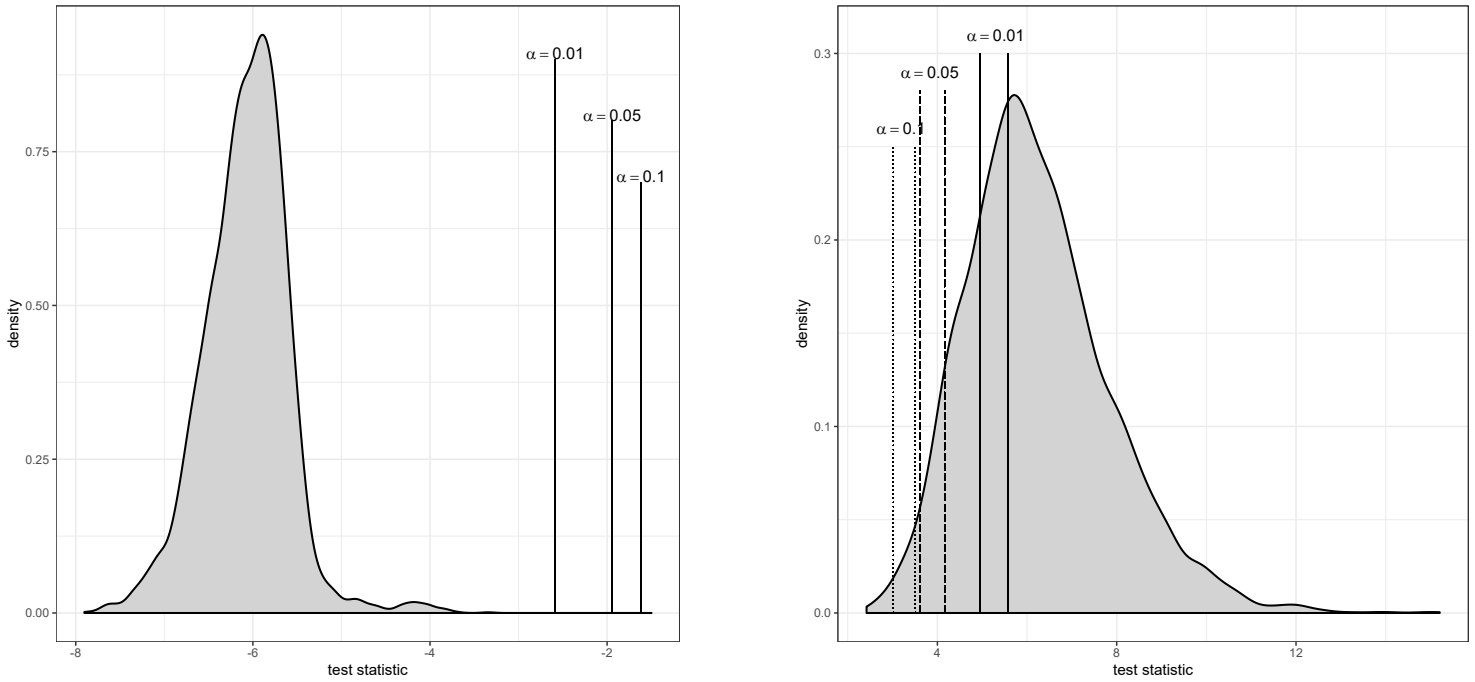
Table 3.3: Percentage of rejection of H_0 when misspecification is at the end of the sample

	Significance level		
Misspecification percentage	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
10%	4.22%	9.7%	14.18%
20%	3.46%	8.26%	12.64%
50%	4.5%	11.6%	17.24%
no misspecification	99.99%	100%	100%

With a higher bias in the long-run coefficients, it is natural to see a decrease in the power of the test. When misdating is at the end of the true break period, the bounds test has low power even when the misdating percentage is small.

To further illustrate the results from the cointegrating test in the simulation study, Figures 3.11 to 3.16 plot the test statistics and critical values for all 10000 simulations of the residual-based test and bounds test. As mentioned above, the residual-based tests have a 100% rejection. This can be seen in Figures 3.11 to 3.16 where all test-statistics in the simulation lie to the left of the critical values of α. For the bounds test, in Figures 3.11 to 3.16, the result can be interpreted as follows. The closer the upper critical value, associated with α, is to the right-hand side of the density the less power the test has. This is because less of the test-statistic observations reject the null. The closer the upper and lower critical values lie with respect to the peak of the density, the more cases there are for which the result of the bounds test were inconclusive. Similarly, the further to the right of the density the lower critical value lies, the more occurrences of not rejecting the null were observed.

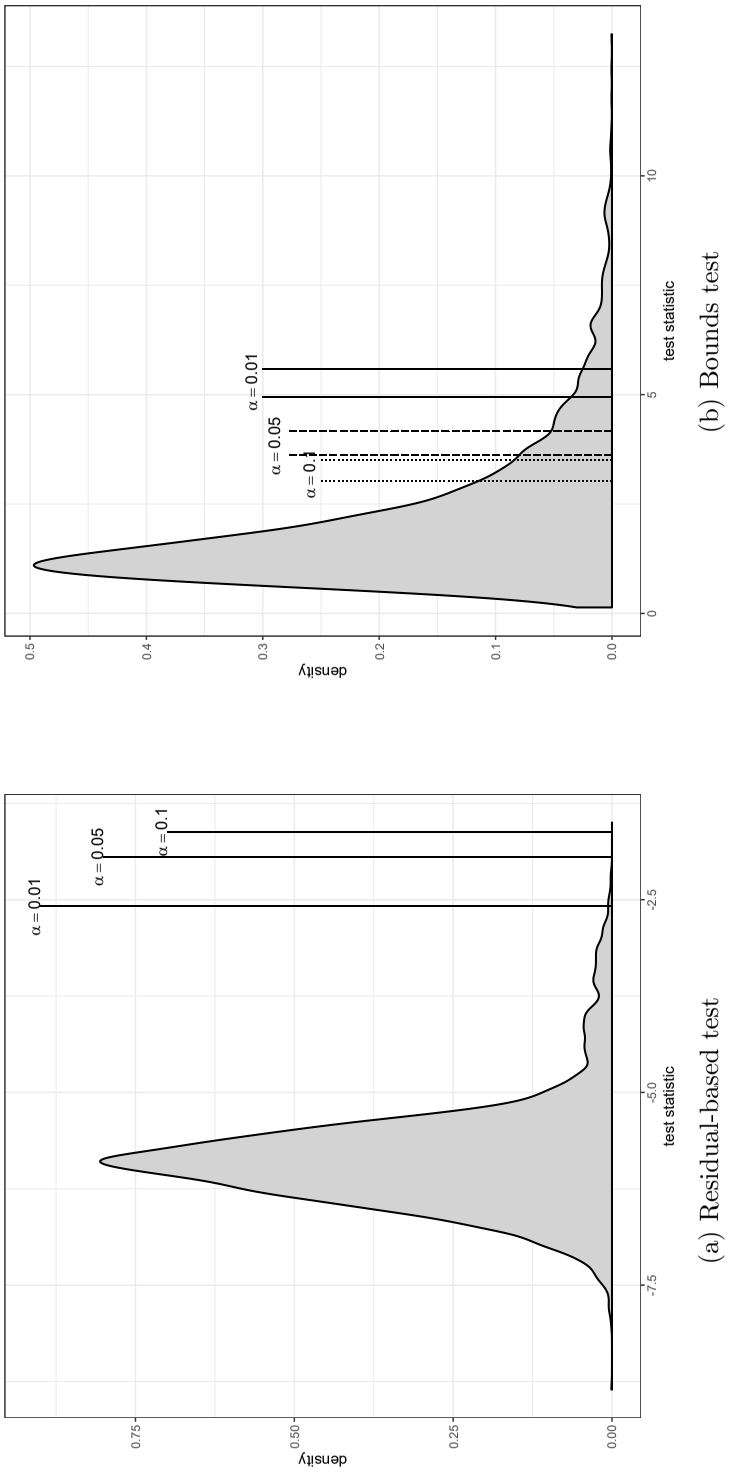
Figures 3.11 to 3.16 corroborate the results reported in Tables 3.2 and 3.3. When the missepcification percentage is small and misdating is at the beginning of the true break period the bounds test has relatively good power (see Figure 3.11 (b)). Conversely, when the misspecification percentage is large or the mis-dating is at the end of the true break period, the bounds test has extremely low power. This can be seen from Figures 3.12 (b) to 3.16 (b), where the upper critical values lie further to the right-hand side of the peak of the density plot compared with Figure 3.11 (b).



(a) Residual-based test

(b) Bounds test

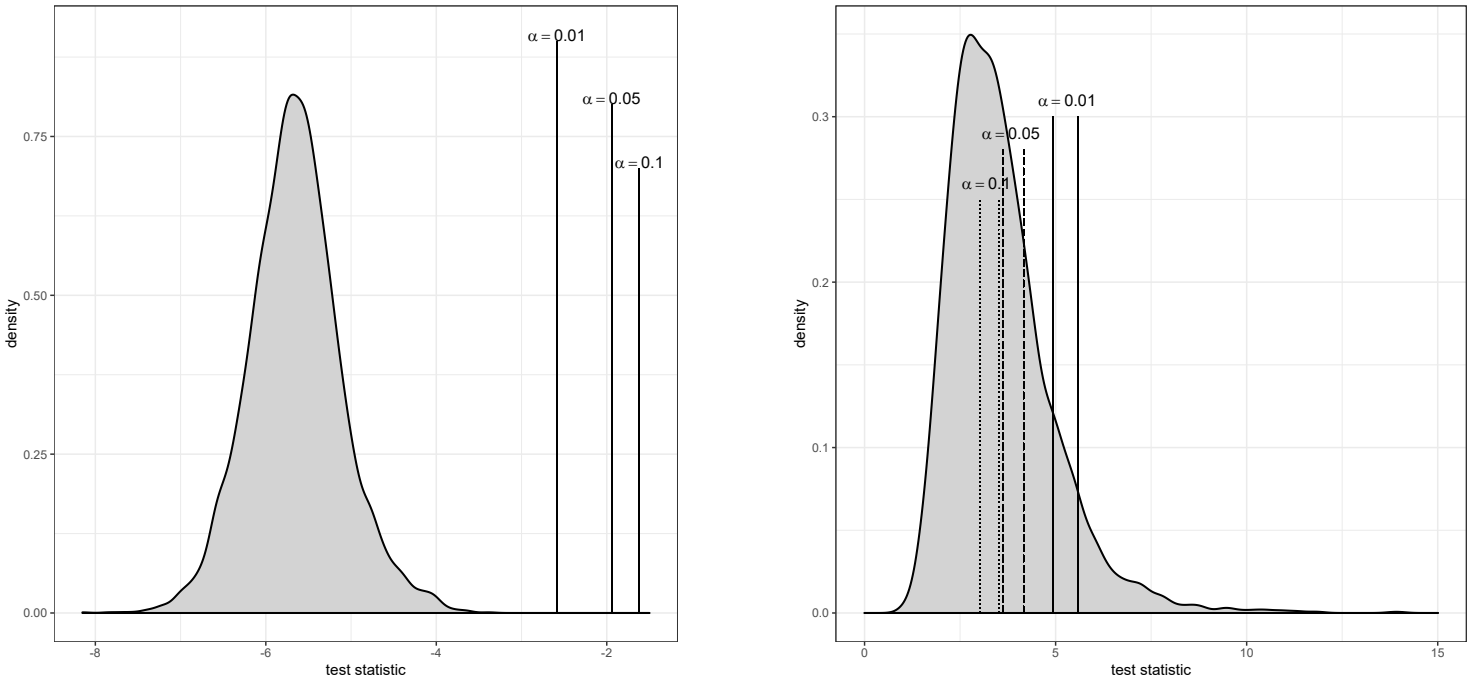
Figure 3.11: $T_m = 10\%$, Misspecification at start of sample



(a) Residual-based test

(b) Bounds test

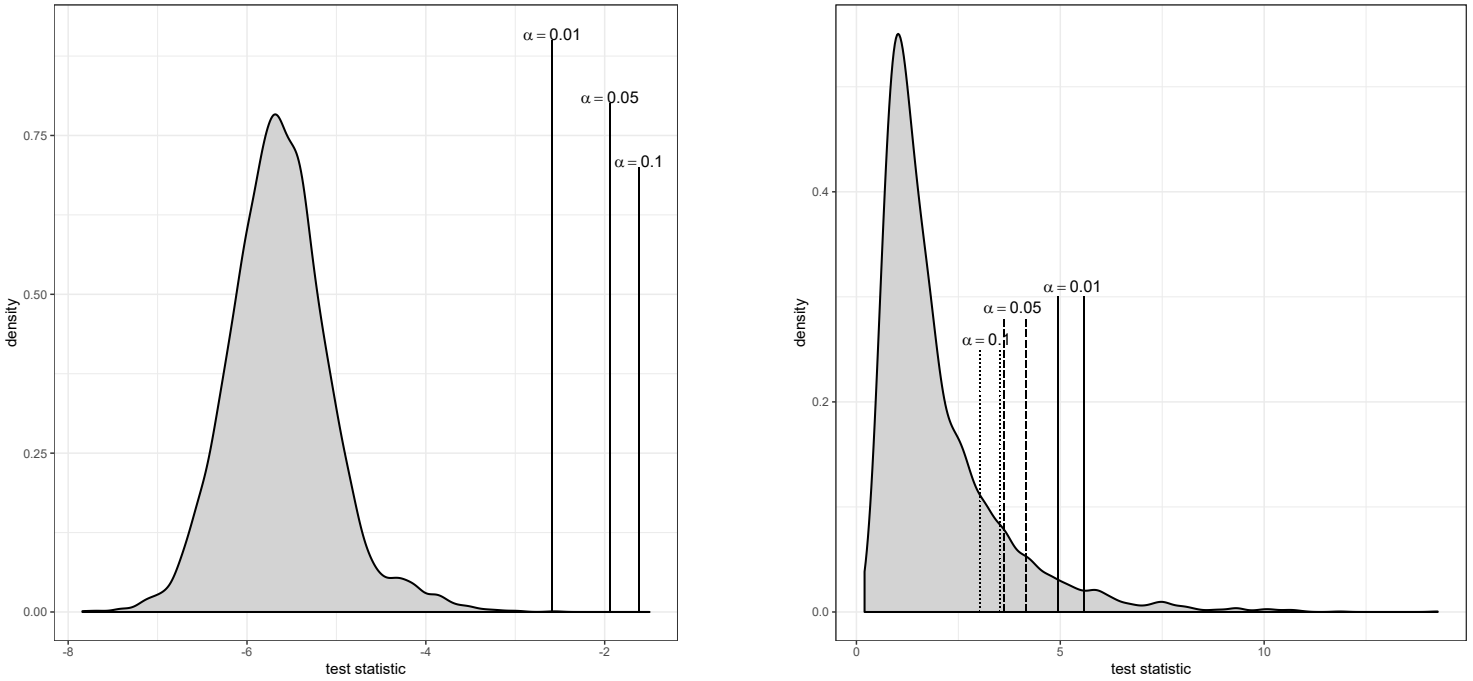
Figure 3.12: $T_m = 10\%$, Misspecification at end of sample



(a) Residual-based test

(b) Bounds test

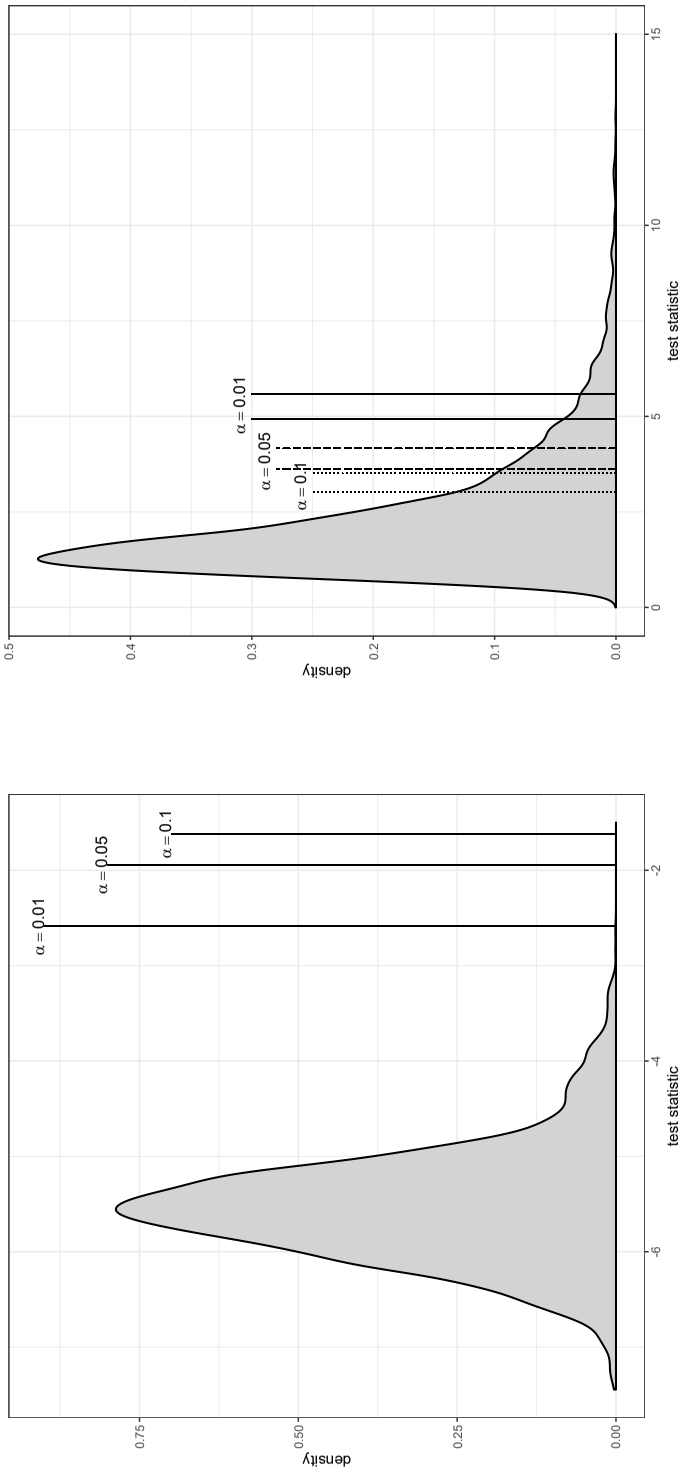
Figure 3.13: $T_m = 20\%$, Misspecification at start of sample



(a) Residual-based test

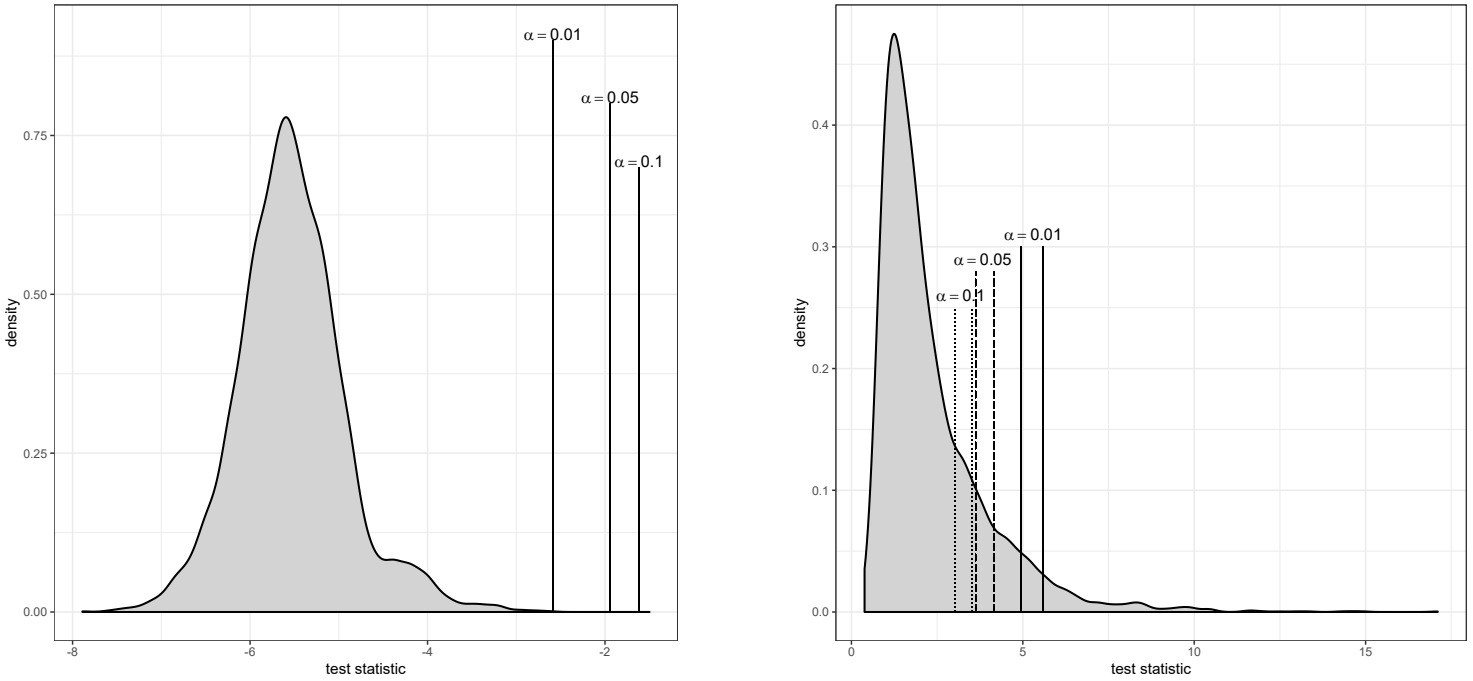
(b) Bounds test

Figure 3.14: $T_m = 20\%$, Misspecification at end of sample



(a) Residual-based test (b) Bounds test

Figure 3.15: $T_m = 50\%$, Misspecification at start of sample



(a) Residualbased test

(b) Bounds test

Figure 3.16: $T_m = 50\%$, Misspecification at end of sample

As shown in Figures 3.11 to 3.16 the residual-based test has very high power in the simulation. While this result is somewhat comforting in terms of testing power for the cointegration test, there is still cause for concern. In this case, the long-run form is not correctly specified, and the coefficient estimates are biased. Even so, the conclusion that would be reached in a practical application would be that the model is well-specified, since a cointegrating relationship exists. Therefore, the result is important to take note of since cointegration tests are unlikely to reject the null of no-cointegration when the dummy variable is misspecified. The residual-based cointegration test should, therefore, be interpreted with caution since it relays no information about the correct specification of the dummy variable. Conversely, the bounds test has extremely low power when the misdating percentage is large or at the end of the true break period. This re-emphasizes the importance of accurate dating, as discussed in Chapter 2, since incorrect conclusions could be reached based on the bounds test when the true break dates are misdated.

3.6.4 Application: Effects of nonstationarity in overcharge estimations

To illustrate how the above results feature in cartel overcharge estimations, the simulations are repeated with the DGP as specified in section 3.5.6. The DGP in section 3.5.6 more closely resembles the properties of overcharge data in practice. In this simulation, there are two regressors representing cost and demand which are both unit root processes. Both these regressors are simulated in a manner where the values are not allowed to be negative since cost and demand drivers cannot contain negative values.

In these simulations, I deliberately misspecified the dummy variable by as much as half of all its observations. Subsequently, I tested whether the t-statistic would still lead to concluding that the coefficient is significant. This was done to illustrate that the t-test will reject the null hypothesis of the coefficient value being 0, despite the fact that the dummy variable is greatly misspecified. Figure 3.17 reports this result.

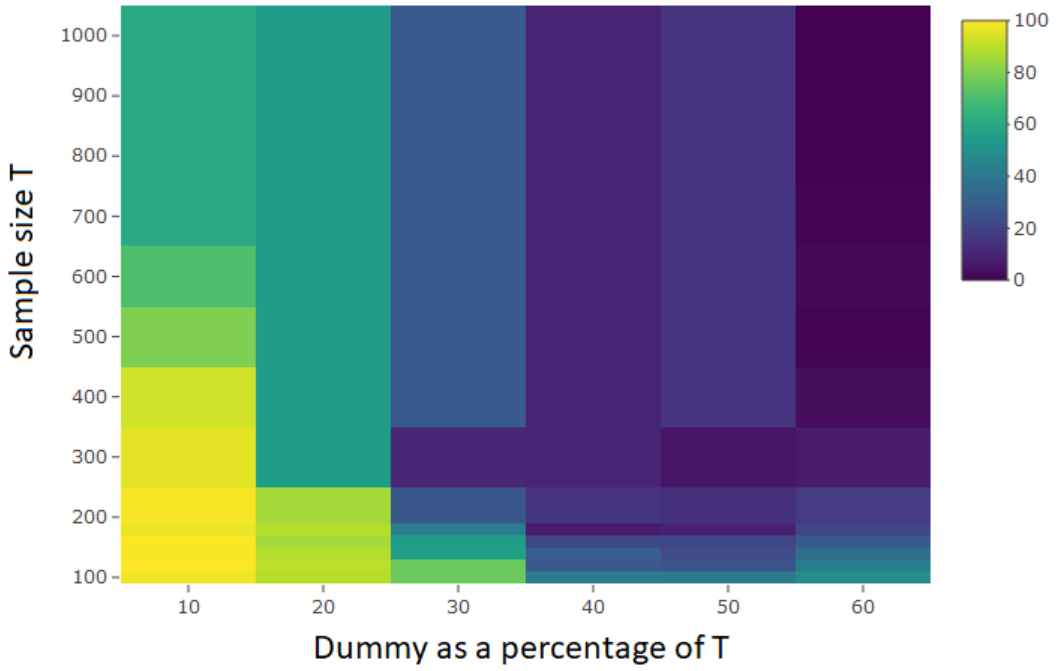


Figure 3.17: Number of H_0 rejections

Conforming with the results in section 3.6.1, the effects are more severe for a small T_c and T. That is, when the collusive period is short in relation to the total sample size there is a high likelihood of misspecifying the collusive period and still concluding that there was a significant effect. Here, the small sample results remain more persistent even when T_c is longer. This is because the t-test tends to over-reject the null hypothesis, as was shown in section 3.6.2. For this application, it appears that having a small T or T_c greatly increases the probability of an incorrect inference. Again, this result draws attention to the importance of correctly dating the break dates as discussed in Chapter 2.

3.7 Conclusion

coefficients of cointegrated systems are often used to describe the long-run relationships between economic time series. While cointegration implies that the relationships are not spurious, the coefficients are no longer asymptotically normally distributed. Previous research has not considered how these effects translate to dummy variables.

In this chapter, the effects that nonstationary data has on dummy variable coefficients is considered. Three main contributions are made. First, I em-

pirically illustrate the difference between the "true" distribution and Student's t-distribution of dummy variable coefficients in finite samples of cointegrated series. By comparing the distributions, I hope to illuminate the problem from a new angle. By considering the magnitude of difference between the distributions, it becomes simpler to infer the probability of drawing incorrect inference when ignoring the effects. Second, when performing bootstrapping on nonstationary data, an important alteration should be made that restricts the re-sampling to only re-sample from similar regimes. This is done to eliminate uncontrolled outliers that will be part of subsequent regressions which will bias the bootstrap results. Last, by applying my results to overcharge estimations I illustrate the problems of incorrect inferential conclusions in a practical context.

Monte Carlo simulations are performed where I incrementally increase the sample size (T) and length of the dummy variable (T_c). For each of the simulated series, I find the ME bootstrapped distribution of the dummy variable coefficient (α_2), and use the K-S statistic to compare this distribution with a t-distribution. The coefficients of the DGP remain fixed for all the simulations in order to directly compare the effects of T and T_c.

For small values of T and T_c I find that the 'true' distribution of α_2 differs significantly from a Student's t-distribution. The difference between the distributions is also shown to be statistically significant in almost all the simulations with small T and T_c. There seems to be less of an effect for large T or T_c. Interestingly, the magnitude and significance of the K-S statistic decreases faster when T_c is increasing than for increasing T. Therefore, in applied work, where the dummy variable does not span a significant portion of the total sample size, careful consideration should be given to the significance of this coefficient when dealing with nonstationary data.

To illustrate the practical implications – where the results in this chapter can be important – the dummy variable approach from the overcharge literature is considered. To obtain results from a typical DGP that can be found in overcharge cases, the simulations are repeated with a few modifications. The DGP is changed to incorporate two regressors that resemble cost (c_t) and demand (d_t) drivers. A restriction is placed on the random walk processes of c_t and d_t where both variables can be increasing or decreasing, but are strictly positive for all values of $t = 1, ..., T$. This is done to represent that fact that cost and demand variables in practice can never contain negative values. Contrary to the previous results, I deliberately misspecify the dummy variable by as much as 50% of its total length.

For the overcharge simulations, the results are equivalent to the K-S results found previously. When misspecifying the dummy variable by as much as 50% of its total length, there is a high likelihood of incorrectly concluding that the coefficient is significant when inference is drawn using the t-distribution. Given that fines and civil damage claims are sometimes based on this measure, it is important to carefully consider the distributional effects that nonstation-

ary data can have. By ignoring this effect, the variable can unknowingly be misspecified.

Cointegration modelling is often considered as a solution to the problems associated with unit roots. However, other problems emerge in cointegration tests when the structural break dates are not accurately captured by the dummy variable. The bounds test has extremely low power when the misspecification of the dummy variable is large and at the start of the true break period. When the misdating is at the end of the true break period, the bounds test has extremely low power regardless of the extent of misdating. Conversely, the residual-based cointegration test has high power (relative to the alternative of cointegration), regardless of the extent of misdating. Even so, caution should be exercised when interpreting the residual-based cointegration test. The residual-based test does not relay any evidence of correct specification of the dummy variable in the long-run equilibrium.

Chapter 4

Practical application

4.1 Introduction

Monetary penalties serves as one of the main deterrents for cartel formation and provide reparation to consumers who suffered welfare losses due to collusion. It is, therefore, important to have accurate techniques to measure the extent to which a cartel has elevated prices. In chapter 2 and 3 I discussed the importance of cartel dating methods and the effects that unit roots can have on overcharge estimates. These are issues that often receive less attention from both practitioners and the literature on cartel damages. In this chapter, I show how some of the issues raised in chapter 2 and 3 translate into practice by using data from the European Sodium Chlorate cartel. It is important to add such an empirical contribution: in practice, the true start and end dates of a cartel may not be known or it may be uncertain. In addition, practitioners rarely consider the non-stationarity of data.

To illustrate the practical implications of the MC simulation results in chapter 2, I apply each of the structural break tests and regime-switching models to the European sodium chlorate cartel. The purpose of this application is to show how different approaches to dating may lead to different conclusions with regard to the nature and size of the overcharge. While in practice, the *true* overcharge is unknown, I draw comparisons between the approaches and test the rationality of the results against *a priori* information about the cartel.

This chapter also considers the practical consequences of dealing with non-stationary data. In practice, the *true* (or empirical) parameter distributions are not known. Therefore, it is not possible to evaluate the difference between the *true* distribution and Student's t-distribution. The particular case data does not contain unit roots, which implies that there is no need to consider conintegration modelling. Nevertheless, I provide an example of the complexities of dealing with unit roots in the presence of breaks. Additionally, a theoretical discussion of possible cointegration modelling techniques is also provided.

The rest of this chapter is organized as follows. In section 4.2 I provide

background and discuss the characteristics of the European sodium chlorate cartel. Section 4.3 applies the modelling frameworks discussed in chapter 2. Section 4.4 provides practical guidance on how to deal with unit roots and section 4.5 concludes.

4.2 The European sodium chlorate case

The industrial production of sodium chlorate involves electrolysis of a sodium chloride solution. Commercially, the main use of sodium chlorate is in the production of chlorine dioxide. Roughly 95% of all industrially used chlorine dioxide is used in the bleaching of pulp (Vogt *et al.*, 2000). Bleached pulp is used to manufacture tissue and printing paper which meets the Elemental Chlorine Free (ECF) paper standards set by the EU.

In the early 1990s demand for sodium chlorate began to stagnate in anticipation of the Totally Chlorine Free (TCF) paper standard that would be adopted. The adoption of the TCF in combination with overcapacity as a result of the TCF placed significant downward pressure on the price of sodium chlorate. To protect themselves against lost profits, the main sodium chlorate producers formed a cartel to stabilize prices in the market. The European Commission (EC) found documentary evidence that the cartel first held meetings on 21 September 1994. In their first meeting AkzoNobel[1] and Kemira agreed on a series of upward price adjustments starting from 1995. From 1995 to 1996 the cartel expanded its members by including smaller suppliers. At its peak, the cartel members were jointly responsible for over 90% of the total sodium chlorate market in Europe.

The cartel was uncovered in March 2003 when AkzoNobel brought the existence of the cartel to the EC's attention and successfully applied for leniency. The EC, in June 2008, found four groups of chemical producers guilty of collusion and price fixing and imposed total fines amounting to 79 million euros[2]. Based on documentary evidence, the Commission established that the cartel formally ended on 9 February 2000 when AkzoNobel had decided to no longer participate.

4.3 Application of structural break tests and regime-switching models

To calculate the overcharge, I rely on the monthly price per ton of sodium chlorate. Figure 4.1 shows the log of the volume-weighted average delivered price of sodium chlorate from January 1993 to December 2005. The price

[1]formerly known as Akzo whose parent company is Eka.

[2]European Commission decision of 11/06/2008 in Case COMP/38.695 – *Sodium Chlorate.*

series is constructed from several large customers of the cartel that make up 50% of the total demand for sodium chlorate in Europe. Supply agreements are usually formed on medium- to long-term contracts and price movements, therefore, represent contractual revisions that are made throughout the year.

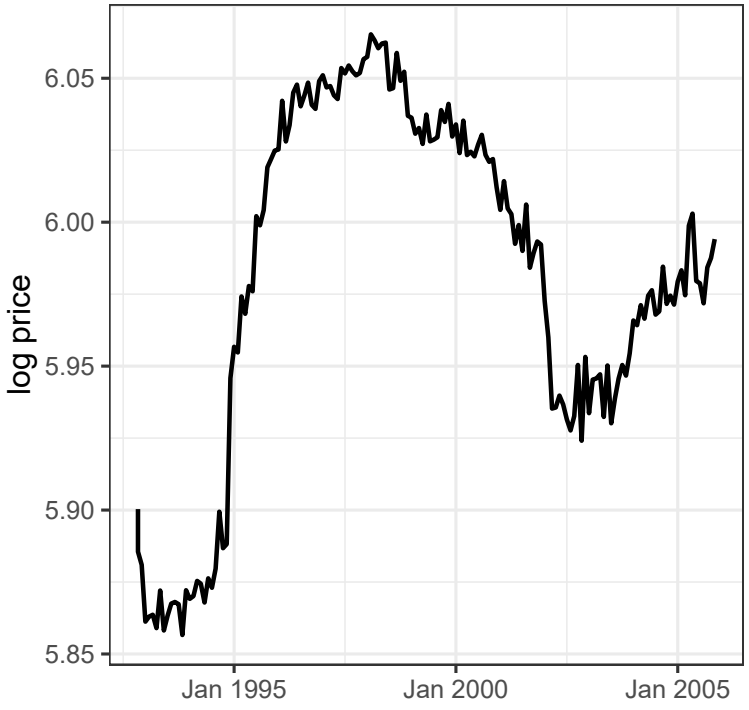


Figure 4.1: Log price of sodium chlorate

To perform the overcharge estimation I rely on the main cost and demand drivers of sodium chlorate for which data is available. The main cost drivers that I control for are electricity prices and labour costs while the main demand factors are European pulp production and production capacity. I specify the following dynamic price equation:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 p_{t-1} + \beta_6 p_{t-2} + \epsilon_t \quad (4.1)$$

where D_t is a dummy variable with $D_t = 1$ when the cartel was effective and $D_t = 0$ elsewhere, x_1 is industrial electricity prices, x_2 is labour costs, x_3 is European pulp production and x_4 is the European production capacity of sodium chlorate[3]. The lagged price regressors, p_{t-1} and p_{t-2}, are included

[3]This data was provided to us by the authors of Boswijk *et al.* (2019) who cite the following sources: Eurostat, Electricity - industrial consumers - half-yearly prices"; OECD System of Unit Labour Cost Indicators; CEPI, European Chemical Pulp Production, 1995Q1 - 2008Q3; Harriman Chemsult, Chemicals Economics Handbook SRI Consulting and CMAI.

to control for short-run dynamics. In the subsequent regressions I measure all the variables' logarithms except for the dummy variable. The regression coefficients can therefore be interpreted as semi-elasticities.

The functional form of equation 4.1 is used to perform the various structural break tests. First, I estimate the model using the formal cartel dates as established by the EC. I then perform the BP, CUSUM, MOSUM and sequential F-tests and use the test results to construct a dummy variable which is used in the subsequent regressions. I also use the MS and TVP models to estimate the overcharge and compare results of the various approaches. Diagnostic tests for each of the estimations are reported in appendix F.

Boswijk *et al.* (2019) compares the results of the formal cartel dates with those obtained by using the Bai-Perron test using the same data as I have. The results of the formal cartel dates compared with the Bai-Perron estimates is therefore a replication of the work done in Boswijk *et al.* (2019). For this comparison I find the same results as presented in Boswijk *et al.* (2019).

Using the formal dates as indicated by the EC (September 1994 – February 2000) to encode the dummy variable I obtain the following estimates:

$$\begin{aligned}\hat{p}_t = {} & 0.78 + 0.013 D_t + 0.007 x_{1t} + 0.023 x_{2t} - 0.005 x_{3t} - 0.054 x_{4t} + 0.596 p_{t-1} + 0.299 p_{t-2} \\ & (0.36)\ (0.00)\quad (0.01)\quad (0.02)\quad (0.01)\quad (0.06)\quad (0.07)\quad (0.07)\end{aligned}$$

The BP test results are reported in Table 4.1. The test results indicate that two break points are the most optimal and the break points have the corresponding break dates as January 1995 and February 2002.

Table 4.1: BP break-test results

supF(1)	1.36
supF(2)	27.34
supF(3)	22.81
supF(4)	21.01
supF(5)	20.16
break dates	Jan 1995
	Feb 2002

To ensure that a single equation specification is appropriate, I estimate a vector autoregressive (VAR) model using the variables from equation 4.1. A dummy variable is included in the VAR using the BP break-test results to ensure that the structural break is controlled for. I then perform a Granger causality test (Granger, 1969). I test weather prices jointly influence the exogenous variables or whether there is evidence of reverse causality in a bivariate model. The results are reported in table 4.2

Table 4.2: Granger causality test

H_0	Test statistic	p-value
p_t does not Granger-cause x_1	1.5884	0.201
p_t does not Granger-cause x_2	0.4167	0.6596
p_t does not Granger-cause x_3	0.0848	0.9187
p_t does not Granger-cause x_4	0.9401	0.4405
p_t does not Granger-cause x_1, x_2	1.0407	0.3857
p_t does not Granger-cause x_1, x_2, x_3	0.8604	0.5238
p_t does not Granger-cause x_1, x_2, x_3, x_4	1.0576	0.3937

The results in table 4.2 indicate that a single equation framework is appropriate. While theoretically one would expect reverse causality between prices and demand, this is often not the case when using practical data[4].

Using the BP break-test results to encode the dummy variable and estimating a regression with functional form as in 4.1 I obtain the following estimates:

$$\begin{aligned}\hat{p}_t = &\underset{(0.37)}{2.033} + \underset{(0.00)}{0.029}D_t + \underset{(0.01)}{0.056}x_{1t} + \underset{(0.01)}{0.031}x_{2t} - \underset{(0.01)}{0.002}x_{3t} - \underset{(0.05)}{0.154}x_{4t} + \underset{(0.07)}{0.455}p_{t-1} + \underset{(0.06)}{0.3}p_{t-2}\end{aligned}$$

Using the BP tests, I find overcharge estimates approximately twice the size compared with using the EC determined dates.

The empirical fluctuation process (efp) of the CUSUM procedure along with the significance bands are illustrated in Figure 4.2. While there is an increase in the epf around 1995, the process does not cross the significance bands. Therefore, when using the CUSUM-based test the conclusion is that there are no structural breaks in the DGP. Given the fact that the data contains a known cartel that admitted to increasing the mean price it is highly unlikely that the true DGP does not contain any breaks. This finding corroborates the MC results presented in section 2.5 which show that the CUSUM-based test performs poorly when there is a mean shift in the DGP. Since the test result implies that there is no break, I do not perform the subsequent regression for this case.

Figure 4.3 reports the results for the efp of the MOSUM-based test with its significance bands. Contrary to the CUSUM approach the MOSUM has an efp that crosses the significance bands at the start of the sample and then again at December 1994. While this test can help establish the start date of the cartel it does not provide any evidence of the effective end date.

While the CUSUM- and MOSUM-based tests are not particularly accurate or helpful at dating the effective start and end dates of the cartel, the results are still helpful. The epf function deviates from 0 around the time of the

[4]See, for example Boshoff and Van Jaarsveld (2019) that uses data from a different case study and arrive at a similar conclusion when testing for weak exogeneity

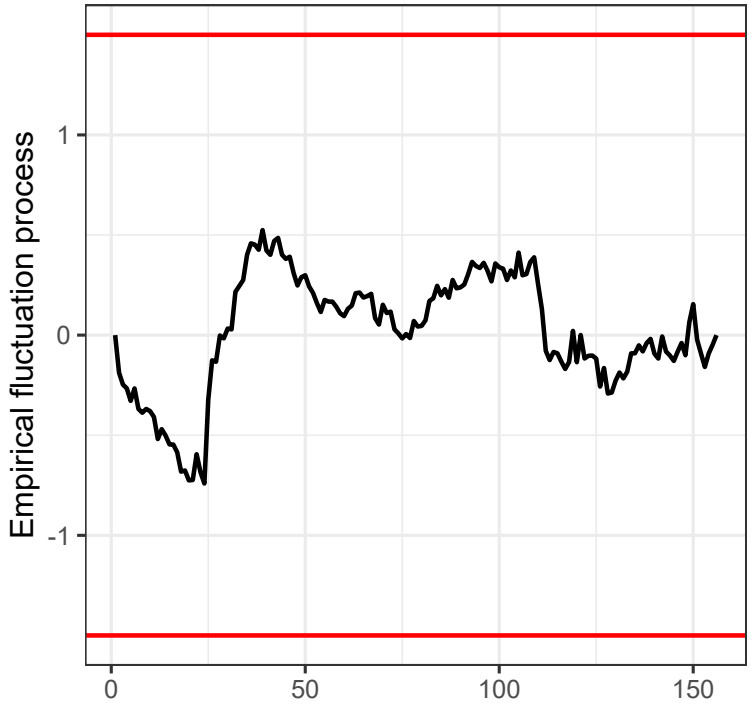


Figure 4.2: CUSUM epf with bounds

effective cartel dates as indicated by the BP test. Additionally, these tests can provide some preliminary statistical evidence that a cartel is present in the market as described by Crede (2019). The use of these two test approaches is, however, not recommended to be used in isolation when attempting to date the start and end dates of the cartel.

Figure 4.4 illustrates the sequential F-statistic results. The tests statistic crosses the significance bound at three points: specifically, at July 1996, January 2000, and April 2000. Table 4.3 reports the estimation result when using the functional form of 4.1 and the sequential F-test results to encode the dummy variable.

When using the sequential F-test to determine the effective cartel date, the resulting overcharge estimation in this case results in an underestimation of what is perceived to be the true effect. The result is similar to using the formal EC determined dates.

Table 4.4 reports the estimation results of the MS model. As previously mentioned the MS model provides an approach that allows for simultaneous estimation of the overcharge and dating of the cartel periods.

When using the MS model to determine the overcharge, the resulting estimate is 0.07 $(10.16 - 10.09)$. This is almost twice as high as the BP result and almost seven times as high as the estimate using the formal cartel dates. Aside from the difference in estimation procedures, the difference between the

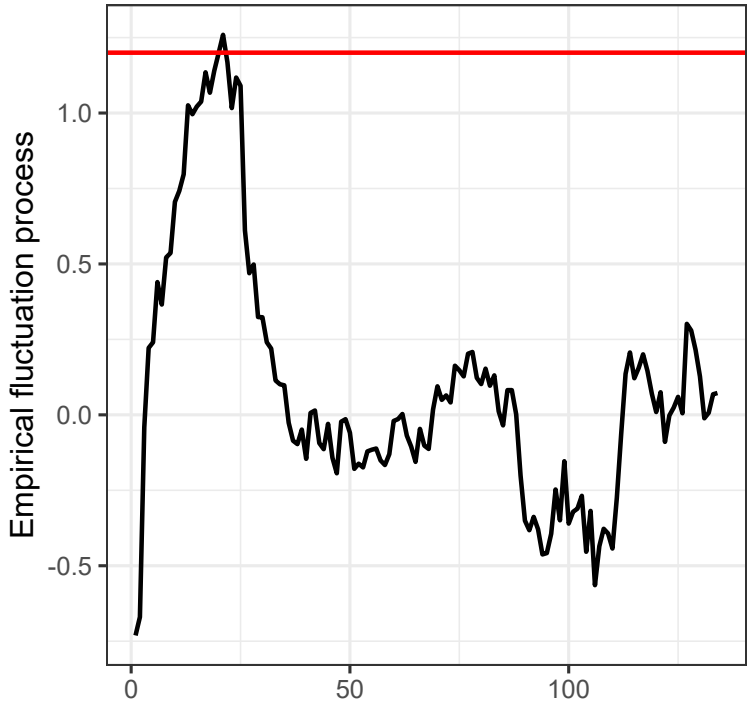


Figure 4.3: MOSUM epf with bounds

Table 4.3: F-test encoded dummy estimates

	Estimate	Std error	t-value
α_1	1.22	0.43	2.83
α_2	0.011	0.48	2.35
x_1	0.031	0.01	2.54
x_2	0.009	0.01	0.01
x_3	0.009	0.02	0.48
x_4	-0.17	0.06	-1.91
p_{t-1}	0.635	0.08	7.99
p_{t-2}	0.241	0.08	3.09
R^2			0.97
Adjusted R^2			0.97

results is mainly driven by the MS model's approach to identifying the regimes. Table 4.5 reports the constant regime transition probabilities. Figure 4.6 illustrates the smoothed and filtered transition probabilities of regime 1 and 2 while Figure 4.5 illustrates the regime realization with the price data.

From the regime realizations in Figure 4.5 it is clear that the MS model incorporates the higher prices towards the end of the sample as part of the collusive price regime. This results in a much higher estimate, since the high

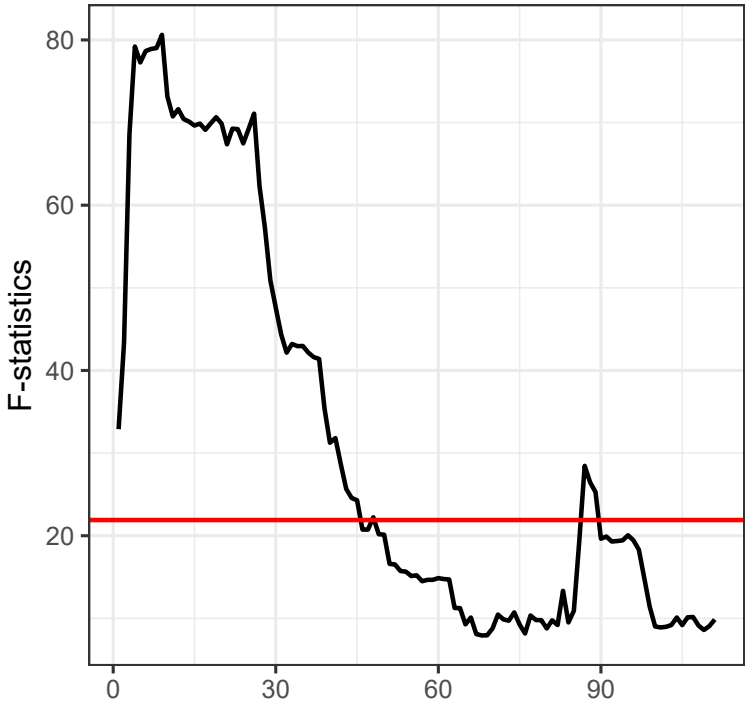


Figure 4.4: F-test with bounds

price at the end of the sample would raise the average competitive price for the other estimates, resulting in underestimation of the effects. It may be challenging to justify the inclusion of these prices as "collusive" in court. The result nevertheless raises interesting questions surrounding lingering effects. When the high price at the end of the sample is removed, using the BP-determined dummy variable yields the same result as that of the MS model. While a court is unlikely to accept the inclusion of this period as collusive, the result suggests that it would be prudent not to treat this data as non-collusive data in the empirical models.

Figure 4.7 shows the intercept of the TVP model. As expected, the TVP model intercept reflects the moving average type behaviour in the price series. As previously mentioned, I argue that – while this model does not explicitly provide the cartel dates – it can still be used to estimate overcharges. The TVP model controls for the determinants of price and the shift in intercept can therefore be interpreted as the change that was driven by the cartel. When using the formal dates, as indicated by the EC, and subtracting the mean of the time-varying intercept from the mean of the rest of the sample the increased log price is equal to 0.061. This estimate is higher than the underestimated price when using the EC dates. Consistent with the results in section 2.5, this estimate is higher than that produced by the BP approach.

Table 4.4: Markov-switching estimates

Regime 1			
	Estimate	Std error	t-value
c_{S_t}	10.09	0.03	328.63
x1	-0.21	0.03	-7.03
x2	0.12	0.02	5.67
x3	-0.08	0.02	-3.73
x4	-1.18	0.02	-63.84
p_{t-1}	0.5	0.06	8.33
p_{t-2}	0.25	0.05	5
Regime 2			
c_{S_t}	10.16	0.03	361.39
x1	-0.21	0.03	-7.03
x2	0.12	0.02	5.67
x3	-0.08	0.02	-3.73
x4	-1.18	0.02	-63.84
p_{t-1}	0.5	0.06	8.33
p_{t-2}	0.25	0.05	5
Adjusted R^2			0.84

Table 4.5: Markov-switching fixed transition probabilities

	Regime 1	Regime 2
Regime 1	0.97	0.02
Regime 2	0.03	0.98

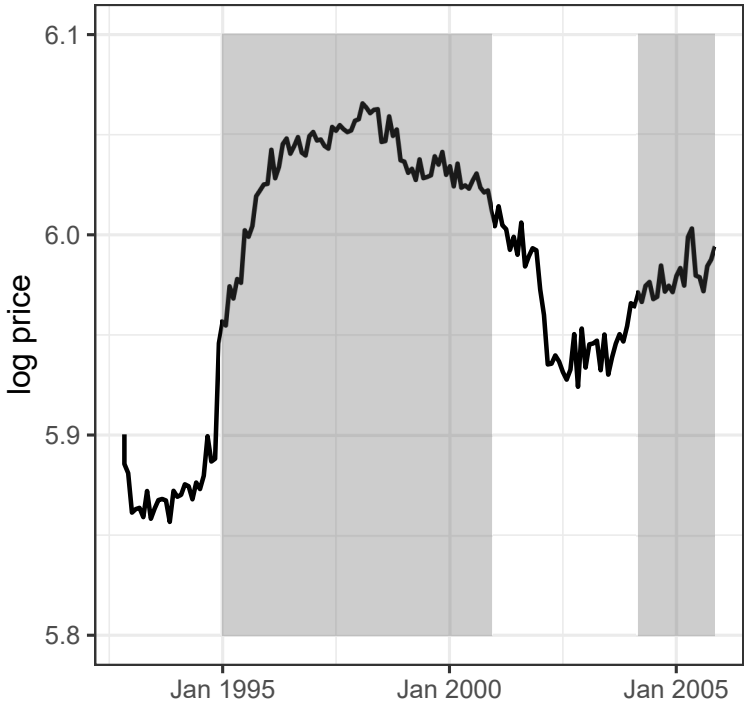
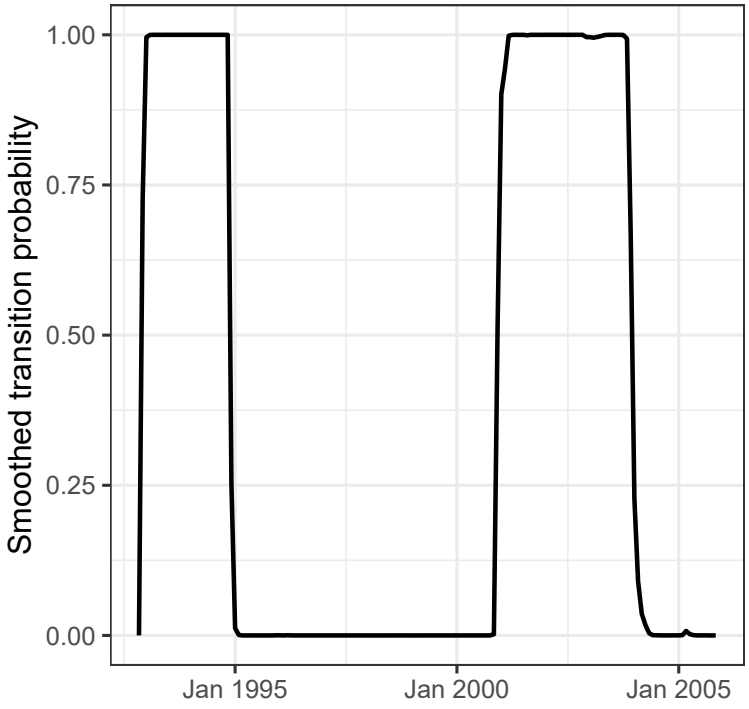
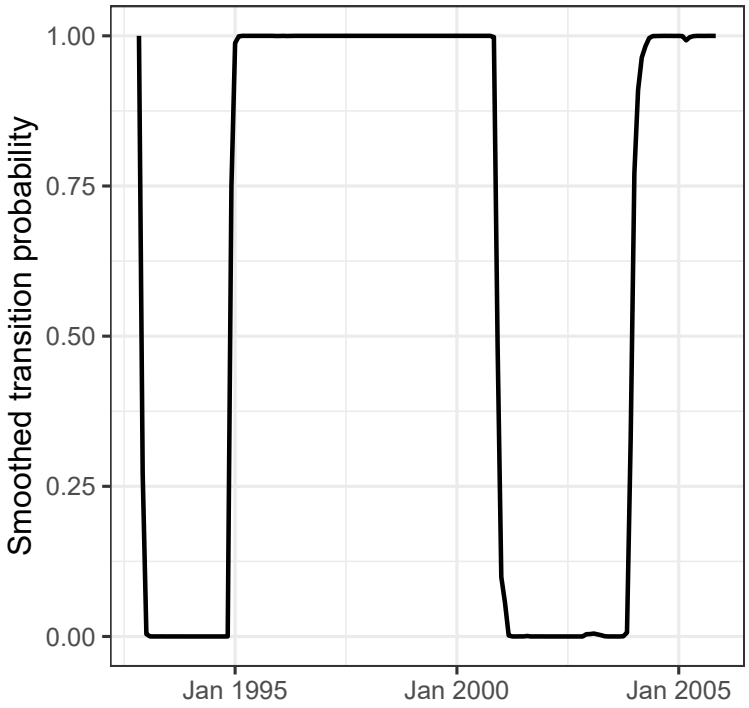


Figure 4.5: Regime realizations with price data



(a) Regime 1



(b) Regime 2

Figure 4.6: Regime probabilities from MS fit

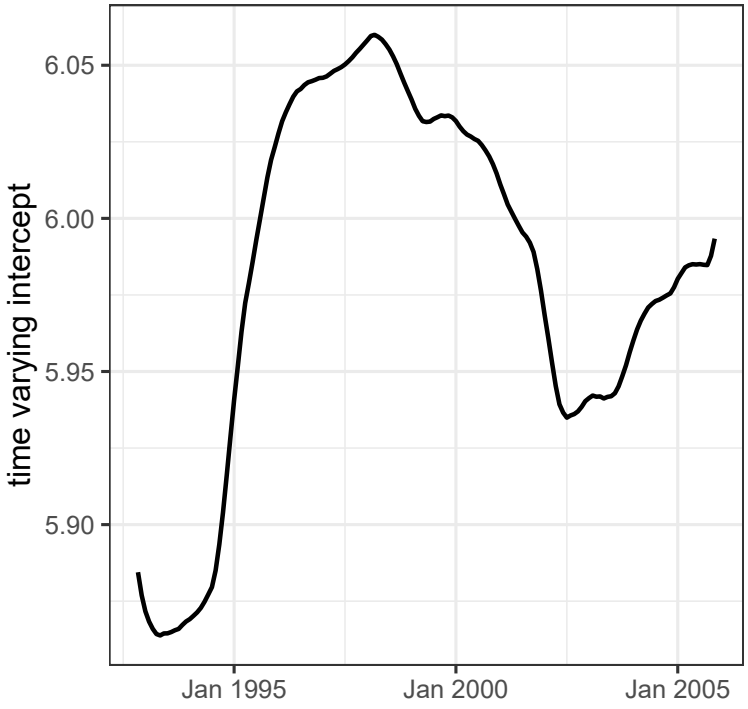


Figure 4.7: TVP model parameter

In summary, the different results obtained from the various approaches illustrate the practical difficulties in determining the correct effective cartel dates and subsequently estimating the overcharge. Different approaches can lead to different conclusions on the length and size of the cartel effect. It is, therefore, important to apply a wide variety of approaches and rationalize these results with the documentary evidence of the case.

4.4 Practical guidance when dealing with unit roots

Considering the effects that nonstationary data can have on dummy variable coefficients is important for applied work. As shown in section 3.6, ignoring the effects can lead to incorrect conclusions and biased estimates. In this section, I outline some considerations and methods to guide applied work.

A simple way to deal with unit roots is to transform the data into first differences. However, this is not a generally appropriate solution. Therefore, it is important to consider an appropriate modelling strategy that can account for the distributional effects of nonstationary data.

The first step is to check if the data is nonstationary by testing for unit roots. Given the nature of the data, specifically, the presence of different

regimes or intercept shifts – this normally straightforward step requires extra care. After unit root testing, a decision has to be made concerning the appropriate modelling strategy. Section 4.4.2 provides a brief discussion in this regard.

4.4.1 Testing for unit roots

The presence of structural breaks in data is a key consideration when testing for unit roots. In the context of cartel damages, data from a collusive market will almost certainly contain structural breaks. That is, the data will exhibit abrupt and/or unexpected changes that cannot be attributed to the independent variables. In fact, this change is exactly what the dummy variable approach aims to isolate: what is the mean change in price after controlling for cost or demand drivers? This is important not only when testing for unit roots in the price variable itself, but also when testing for unit roots in the cost and demand variables. For example, a collusive agreement to restrict output may affect the cost of production, implying a structural break. In turn, abrupt changes in price, cost and output can also cause structural breaks in the demand variable. When conducting unit root tests on cartel data, it is therefore important to note that accompanying structural breaks may translate into reduced statistical power.

Perron (1989, 1990) shows that the autoregressive coefficient is asymptotically biased towards one when performing the augmented Dickey-Fuller unit root test in which allowance is made for when there is a structural change in the trend function. If there is a structural change in the slope, the autoregressive parameter's probability limit is exactly equal to one. This results in a substantial loss in power since the null hypothesis of a unit root is less likely to be rejected, even when no unit root is present. This issue is not specific to the augmented Dickey-Fuller unit root test. Lee *et al.* (1997) show that the presence of structural breaks also affects the Q or KPSS test, by inducing size distortions and leading these tests to reject the null hypothesis too often. Hence, for any form of unit root testing, structural breaks introduces a bias towards favouring the unit root hypothesis.

An appropriate unit root test for cartel data should meet two criteria. First, the test must allow for endogenous determination of the break dates. As discussed in chapter 2, the specific break dates are not always known and should be determined empirically. Second, the test must allow for the appropriate number of breaks depending on the data. Tests that allow for only a single endogenous break, such as Perron (1997) and Zivot and Andrews (1992), may not be appropriate when the data contains more than one structural break. For data – such as presented in this chapter – where the DGP contains two structural breaks, the Lee and Strazicich (2003) unit root test. The Lee-Strazicich unit root test can accommodate two structural breaks, both under the null and

alternative hypothesis. Therefore, rejection of the null unambiguously implies stationarity.

To illustrate the complexities of unit root testing in the case of cartel data, table 4.6 shows the results of the Dickey and Fuller (1981) and Kwiatkowski *et al.* (1992) tests. Table 4.7 report the results of the Perron (1997) and Zivot and Andrews (1992) tests.

Table 4.6: Standard unit root tests

	ADF (none)	ADF (constant)	ADF (trend & constant)	KPSS (level-stationary)	KPSS (trend-stationary)
p_t	0.985	-1.532	-1.252	0.274***	0.046***
x_1	-0.552	-5.560***	-6.688***	0.248***	0.231
x_2	3.802	-6.07***	-5.932***	0.234***	0.078***
x_3	0.843	-5.905***	-5.501***	0.185***	0.032***
x_4	-1.205	-4.628***	-4.228***	0.257***	0.211

Table 4.7: Unit root tests with a single break

	PP (constant)	PP (trend)	ZA (intercept)	ZA (trend)	ZA (trend & intercept)
p_t	-1.623	-1.451	-1.199	-3.648	-0.115
x_1	-4.225***	-6.315***	-7.011***	-3.964	-6.733***
x_2	-5.388***	-5.312***	-6.05***	-5.628***	-6.19***
x_3	-4.353***	-4.666***	-7.367***	-6.993***	-5.49***
x_4	-4.314***	-3.927**	-4.206	-6.69***	-5.814***

Unsurprisingly, the standard unit root tests in table 4.6 indicate that a unit root is present in the price series, p_t. For all of the other determinants – which do not contain the same structural break that is in the price series – there is clear evidence that no unit roots are present. When allowing for only a single break, both the Perron (1997) and Zivot and Andrews (1992) fail to reject the null of a unit root. When performing the Lee and Strazicich (2003) test I obtain a test statistic of -5.775 which implies rejection of the null hypothesis in favour of the alternative indicating that the price series, p_t contains two level shifts and no unit roots.

4.4.2 Modelling strategy

Given the inferential consequences of non-stationary data, discussed in chapter 3, there has been substantial development in procedures that provide asymptotically efficient results for cointegrating vectors. The two most frequently

used methods in applied work is dynamic OLS (DOLS) (Stock and Watson, 1993) and fully modified OLS (FMLS) (Phillips and Hansen, 1990). Both methods take care of the small endogeneity and small sample bias by including the leads and lags of the first differenced regressors.

The DOLS regression model estimates a regression of the following form:

$$y_t = \alpha + \beta x_t + \sum_{j=-q}^{p} \omega_j \Delta x_{t-j} + u_{t1} \tag{4.2}$$
$$u_{t1} \sim IN(0, \sigma_\varepsilon^2)$$

where q is the number of leads and p is the number of lags. As a result of the added leads and lags of x_t, the β parameter will be asymptotically normally distributed and efficient.

The FMLS model is a modification of the OLS estimator where a semi-parametric correction is made, that removes the regressor endogeneity and asymptotic bias terms. FMLS is performed by writing equation 4.3 in the triangular representation:

$$y_t = \theta x_t + u_{t1} \tag{4.3}$$
$$\Delta x_t = -\psi(y_{t-1} - \theta x_{t-1}) + \epsilon_{t2} = u_{t2} \tag{4.4}$$

Let $\hat{\Sigma}$ and $\hat{\Lambda}$ be consistent estimators of Σ and Λ which is the long run covariance matrix and one-sided long run covariance matrix of u_t respectively. Then the FMLS estimator is given by:

$$\hat{\theta}_{FMLS} = \left(\sum_{t=1}^{T} x_t^2 \right)^{-1} \left(\sum_{t=1}^{T} x_t y_t^+ - T\hat{J}^+ \right) \tag{4.5}$$

where the two terms $y_t^+ = y_t \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \Delta x_t$ and $\hat{J}^+ = \hat{\Lambda}_{21} - \hat{\Lambda}_{22} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$ will remove the regressor endogeneity and asymptotic bias.

As with any econometric model, these methods need to be applied with caution. FMLS imposes the additional requirements that all variables should be integrated of the same order and that the regressors themselves should not be cointegrated. It is for this reason that DOLS is often preferred in applied work. However, with DOLS, including leads and lags of the regressors can induce multicollinearity.

When working with finite samples, DOLS and FMLS are also not entirely flawless in terms of parameter inference. Through a Monte Carlo simulation study, Cappuccio and Lubian (2001) show that the parameter efficiency of DOLS and FMLS worsens substantially when the unit root process is borderline stationary. Abnormally high standard deviations are also reported when there is a high signal-to-noise ratio and weak exogeneity. In a similar study, Maki (2011) shows that the estimator's t-statistics suffer from severe size distortions when the cointegration errors have non-linear adjustments and weak exogeneity is present.

An alternative modelling approach to the DOLS and FMLS would be to consider methods that allow for regime-switching cointegration. Regime-switching cointegration is related to the work of Hanson (1992) and Gregory *et al.* (1996) that devised cointegration tests that allows for a structural break in the cointegrating relationship. One method of dealing with cointegrating relationships that are subject to structural breaks is to employ a Markov-switching vector error correction model (MS-VECM). These models are discussed in Krolzig (2013). More recently, MS-VECM have been studied in a Bayesian setting; see for example Jochmann and Koop (2015). There are practical difficulties, including constraints of time and data, that do not often preclude estimating multiple equation systems in damages cases. It is often sufficient to settle on single-equation methods, provided that the weak exogeneity requirements are satisfied. Nevertheless, it would be useful for future research to explore the use of these methods in the context of cartel overcharge estimations.

In summary, there are three methods that can be used to work around the issue of non-normally distributed parameters in cointegrated systems. DOLS and FMLS can be used to obtain valid t-statistics. An alternative approach could be to use simple OLS with ME bootstrapping, as explained in this article. This method might be more robust since the bootstrapping is not conditional on the specific model specification, and will obtain valid critical values irrespective of the sample size.

As discussed in section 3.4, regardless of which method is preferred, the dummy variable needs to be restricted to the long-run equilibrium. If the dummy variable is not restricted it will influence both the stationary and non-stationary components of the model. If the dummy variable is not restricted to the long-run component, the model dynamics imply that prices cannot revert to their competitive level.

4.5 Conclusion

While there are ample methods available for dealing with structural change in the econometric literature, their empirical application in the overcharge estimation literature is limited. In chapter 2, I compared a variety of these methods in different settings. To illustrate how these approaches would translate into practice I apply each of the methods to data from the European Sodium Chlorate cartel. I compare the results with the formally established cartel dates as set out by the EC. I find that the BP approach and MS model produces overcharges that are more than double those obtained when those obtained by using the formal cartel dates. While the other approaches do provide evidence of structural breaks, I find that they are not reliable at correctly determining the effective start and end dates of the cartel.

The MS and TVP model results also alludes to an interesting legal debate.

Even after cartels are prosecuted their previous conduct could alter market prices in the future. This can be due to restricted production investment during the cartel period or previous familiarity with competitors that enables tacit collusion. The pertinent legal question is, therefore, whether cartels should be held liable for higher price levels even after the formal court determined end date.

In chapter 3 I discussed the effects that unit roots can have on dummy variable coefficients and cointegration testing. For the practical case here, the data does not appear to contain any unit roots and as a result I do not illustrate the effects of unit roots in this particular practical case. However, testing for unit roots in the presence of structural breaks is not trivial. I show that incorrect conclusions would be reached if unit root test are used that do not account for the break dates.

Given the inferential consequences of unit roots I provide theoretical guidance in this chapter. In applied work, DOLS and FMLS can be used to remove bias and regressor endogeneity. Alternatively, the ME bootstrap as illustrated here can be applied to obtain robust critical values irrespective of the model used.

Chapter 5

Conclusion

Structural change in econometric models is typically defined as parameter instability. If this is not taken into account, all inference will be biased. Testing the model's fit using goodness-of-fit measures and residual analysis is not helpful in this case. As Breiman *et al.* (2001) notes, many standard tests will indicate a good fit unless the change in parameters is extreme. As a result, a large econometric literature has developed around the testing of structural breaks. Related to this, a great deal of the unit root literature has given attention to the interactions between unit roots and structural breaks. One reason is due to the intricate interplay between the two issues, where one can easily be mistaken for another. This dissertation studies a series of unexplored issues related to tests of structural change and the modelling thereof.

If a time series is subject to structural breaks, there are two modelling approaches available. One approach is to use a structural break test to determine the start and end dates of the break, and subsequently to estimate a model with a dummy variable that controls for the breaks. Alternatively, a regime-switching or time-varying model can be used which accounts for structural change at unknown dates. Since the early work of Quandt (1958, 1960) and Gardner (1969), there have been many tests developed to determine the timing of structural breaks. An area that has received less attention is how these tests influence parameter estimates when used to inform the construction of a dummy variable that controls for the breaks. Additionally, there has been little research on the trade-offs between using a model with a structural break test informed dummy variable and using a regime-switching model. The core of Chapter 2 is dedicated to providing insights into these two matters.

Time series is often described as consisting of a trend and cyclical component. Since the work of Nelson and Plosser (1982), it has become standard practice to investigate whether the trend component can be characterised as a random walk. The typical procedure involves testing for unit roots and subsequently ensuring that the variables in the econometric model are cointegrated. Dealing with inference on parameter estimates in cointegrated regressions is involved. This is due to the fact that the long-run parameters will typically

have non-normal asymptotic distributions. An open question is how these effects translate to dummy variables in cointegrated regressions. Chapter 3 deals with this question. Related to Chapter 2, Chapter 3 also investigates the effects of misdating structural break dates on cointegration tests.

An area where the modelling of structural change is particularly relevant is the detection and punishment of cartels. The focus of the applications in this dissertation is specifically aimed at advancing this literature, but the insights are more generally relevant. This Chapter provides a summary of Chapter 2 and Chapter 3 in sections 5.1 and 5.2 respectively. Some of the limitations of the findings are discussed in section 5.4 and I conclude by providing guidance for future research in section 5.5.

5.1 Cartel dating

Any group of firms who wish to protect their interests can form a cartel. Once formed, members can avoid price competition by artificially fixing prices. Due to the negative effects, cartels and price fixing is prohibited in most jurisdictions. An important instrument in cartel enforcement is damage directives, which provide any parties that suffered heavily due to the cartel, the right to be compensated. In order to receive compensation the plaintiff must first demonstrate and quantify the existence of the damages Rubinfeld (2012). Econometric models have become the de facto technique to determine damages.

To determine the total damage there has to be an estimation of the overcharge. That is, the monetary value that a consumer over-paid for products in the cartelized market. One of the most commonly used methods to estimate overcharge involves the use of benchmarks. The two standard benchmark approaches are the *dummy variable* and the *forecasting* approach. In the dummy variable approach, a regression model is estimated with a dummy variable spanning over the entire collusive period. The forecasting approach estimates a model using data outside of the contravention period, and subsequently uses the estimates to construct a counter-factual price that would have prevailed in the absence of collusion. The main focus of this dissertation is on the dummy variable approach.

A pivotal input in any benchmark approach is the effective start and end date of the cartel. Misdating the start and end date will cause under-estimation of the true overcharge (Boswijk *et al.*, 2019). For some time, the standard practice was to rely on documentary evidence which relates to the first and last communications between colluding firms. However, as shown in many practical applications, these dates may not necessarily coincide with the dates on which the cartel was able to effectively influence prices (Hüschelrath *et al.*, 2016; Boswijk *et al.*, 2019; Boshoff and Van Jaarsveld, 2019; Crede, 2019). To overcome this issue, structural break tests and regime-switching models have

been proposed to statistically determine the start and end dates on which the cartel effectively influenced prices.

An area that requires further understanding is how the size and power properties of structural break tests influence parameter estimates. That is, how using a structural break test to inform the construction of a dummy variable translates into potential parameter bias in subsequent regressions. There has been voluminous research on the size and power properties of various tests (Perron *et al.*, 2006; Casini and Perron, 2019). Interestingly, there has been little research into how these properties would translate to dummy variable parameters when used in a two-step estimation procedure. Furthermore, how these tests compare with one another under various conditions has not been studied in detail. A related question, is how models with structural break date informed dummy variables compare with regime-switching models. In the context of overcharge, it is important to understand how these various approaches perform under various conditions to establish guidelines for future cases. Chapter 2 provides answers to these questions.

To address these questions, I rely on Monte Carlo (MC) simulations. To evaluate the performance of various methods, I need to know the exact form of the sampling distribution and be able to construct it by drawing an infinite number of samples from the population. This can be achieved through MC simulations. Relying on practical data to assess the properties of various test statistics and estimators are not viable. This is because in practice the data does not constitute the entire population, and therefore the sampling distribution cannot be determined with certainty. Furthermore, exact knowledge of the functional form and specification of the data generating process (DGP) is required. Without this knowledge, there is a high probability that the test errors and parameter bias are due to misspecification.

In any overcharge estimation, the implicit assumption is that the cartel increased the mean of the price over the collusive period. Hence, I are interested specifically in the performance of various approaches on different types of mean shifts in the DGP. Using MC simulations, I investigate the relative performance of the various approaches by applying each method to four different DGPs. As discussed in section 2.2, each of the four DGPs are motivated by various case studies and theory of collusive behaviour. The first case is the standard deterministic case, where a break occurs that immediately shifts the mean of the price series at the start date, and immediately returns to the pre-break price levels at the end date. In the second case, the breaks have the same influence on price as in the first case, but are instead driven by a Markov process. The third case is where the breaks are recurrent. In this case, the DGP is subject to breaks where the mean shifts up and down more than once. The fourth case represents the phenomenon where there are transition phases. In this case, the mean takes time to fully shift up and down.

I evaluate the performance of four structural break tests and regime-switching models, namely, CUSUM, MOSUM, sequential F-test, Bai-Perron (BP), Markov-

switching (MS) and a time-varying-parameter model (TVP). I find that in general the MS model performs well in each of the four cases. However, there is a significant amount of variance between the estimates of the MS model for each of the simulated estimations. One of the reasons for this is due to complications associated with numerical optimization procedures which, is discussed in section 2.6. In the first case of a deterministic DGP, and third case of recurrent changes, the BP approach performs the best. However, if the breaks in the DGP are driven by a probability process or contain transition phases, as in the second and fourth case, the BP approach results in significant parameter bias. In all four of the cases, I find that the CUSUM, MOSUM and sequential F-test perform poorly and will result in significant parameter bias. The conclusion is that these three methods are not well suited to deal with the specific breaks in the mean of the DGP.

There are several lessons that policy makers can draw from the evidence presented in chapter 2. First, the Chapter re-emphasize the importance of statistically determining the effective cartel dates. Relying on case evidence alone can result in under-estimation of the true overcharge. Second, I advocate that multiple approaches should be used. Various methods perform better or worse under various conditions. There is no golden standard on which approach will always be best. It is therefore important to use a variety of different methods in tandem with case evidence and robustness checks to determine what the appropriate overcharge estimate is.

5.2 Unit roots and dummy variables

Estimating overcharge is an exercise in estimating the size of a structural break. As previously discussed, when the break dates are known, a dummy variable can be used to control for the mean shift in the series and the dummy variable coefficient will indicate the size of the shift. The dummy variable coefficient is therefore subject to intense scrutiny especially surrounding its statistical significance. It is in the interest of cartel members to argue that their collusion did not have a statistically significant impact on prices. The onus lies with the claimant to prove that this is not the case.

When time series contain unit roots, the inference on statistical significance of the model coefficients becomes troublesome. This is because the coefficient distributions will no longer be asymptotically normal. This presents a particular concern for overcharge estimations where the statistical significance of the dummy variable coefficient is pivotal to successful damage claims. An un-addressed issue is how the presence of unit roots will affect the dummy variables' coefficient distribution. Furthermore, when dealing with unit roots it is standard procedure to test whether a cointegrating relationship exists. An area that requires further insights is how misdating the break dates will influence cointegration test results. In overcharge estimation, this is related to

the issue in Chapter 2 where formal and effective cartel dates do not always coincide. Chapter 3 addresses these two questions.

The asymptotic distributions of the regressors in cointegrating regressions are stochastic due to their dependence on nuisance coefficients. Therefore, deriving the exact distribution of the coefficients is not plausible. Instead, I investigate the difference between the standard t-distribution and that of the dummy variable coefficient in a cointegrated regression. This demonstrates the conditions under which the distributional differences are of particular concern. There has been extensive evidence proving that coefficient estimates in cointegrated regressions do not have asymptotic t-distributions. A valuable contribution in this Chapter is to understand how significant the distortions in distribution can be.

Campos *et al.* (1996) have studied the performance of cointegration tests in the presence of structural breaks. The authors find that structural breaks have little effect on the size of standard cointegration tests. I extend this analysis by considering how the size of cointegration tests is influenced when the break dates are misspecified. Understanding this effect is important for two reasons. First, in practical application there is a risk that the break dates are misspecified and the conclusion is that a cointegrating relationship exists and the estimates are therefore reliable. Second, there is a risk that the null hypothesis of cointegration is rejected, leading the modeller to believe that there is no long-run relationship between the variables, when in fact the test was rejected due to misdating of the break dates.

To determine the differences between the t-distribution and the true distribution dummy variable coefficient in cointegrated regressions, I rely on MC simulations. The difference between the distributions is tested using the Kolmogorov-Smirnov test. I investigate this difference for various sizes of the total sample (T) and length[1] of the dummy variable (T_c). The results show that the differences between the distributions are at their largest for small values of T and T_c. The difference between the distributions decrease as T and T_c increases. Interestingly, I find that increasing T_c has a much greater effect on reducing the difference between the distributions than increasing T.

While important, knowing the difference between the t-distribution and the dummy variable coefficient is not enough to emphasize this issue in practical applications. I therefore evaluate the size distortion of incorrectly using a t-test by constructing p-value plots. I find that incorrectly using a t-test will systematically lead to over-rejection of the null hypothesis. That is, there is a high probability of incorrectly concluding that the dummy variable coefficient is statistically significant when it is not.

Related to Chapter 2, I are particularly concerned with the effects of misdating the break dates. Specifically, in the context of unit roots, I need to

[1]Length in this context refers to the number of consecutive 1s in the dummy variable's construction

consider the effects that misdating the break dates will have on cointegration tests. It should be recalled that typically the dummy variable will be constructed off the basis of documentary evidence or structural break tests. It is therefore important to know if cointegration tests on this model will be influenced by errors in the first step where the break dates were determined. To evaluate these effects, I again rely on MC simulations and consider the cointegration tests of Engle and Granger (1987) and Pesaran *et al.* (2001).

I find that the residual-based test of Engle and Granger (1987) enjoys high power regardless of the degree of misdating. Even so, caution should be exercised when interpreting the residual-based cointegration test. The residual-based test does not relay any evidence of correct specification of the dummy variable in the long-run equilibrium. The bounds test of Pesaran *et al.* (2001) has extremely low power when the misspecification of the dummy variable is large and at the start of the true break period. When the misdating is at the end of the true break period, the bounds test has extremely low power regardless of the extent of misdating. The results from the bounds test re-emphasizes the importance of accurate cartel dating.

5.3 Practical application

To illustrate how the methods from chapter 2 would be used in a practical case, I consider the European *Sodium chlorate cartel*. I compare the estimates of statistically determined structural breaks to the formal breaks as outlined by the European Commission (EC). As expected, I find overcharge estimates more than double when using the BP and MS approach, compared with the EC determined break dates. Using the BP approach, I find estimates that are the same as those reported in Boswijk *et al.* (2019). Consistent with The simulation results, I find that the CUSUM, MOSUM, and F-test approaches do not provide reliable break dates.

To illustrate the complexities of dealing with unit roots, this chapter provides a brief discussion of the relevant literature. Due to the nature of cartel price data, a unit root test that allows for multiple structural changes is required. The results show that not adequately controlling for the structural breaks in the unit root test the incorrect conclusion that the data contains a unit root when it does not is often made. The chapter concludes by providing a discussion of various modelling strategies that account for the issues presented by the presence of unit roots in the data.

5.4 Limitations

As with any research, the results need to be interpreted with caution since they are subject to a specific methodological design. While I went to great lengths

to ensure the scientific validity of the findings there are some constraints that are unavoidable. This section discusses some limitations of the dissertation to better place the results in context.

It is important to note that all the findings in Chapters 2 and 3 are finite sample findings. I do not evaluate the asymptotic properties of structural break tests and inference on dummy variable coefficients. In Chapter 3 I do consider the effects of various sample sizes, but these sample sizes are still restricted. Chapter 2 is more restrictive in this regard. Due to the large number of tests and models investigated, as well as the various forms of structural breaks that are considered, I also chose not to vary the sample size. The choice of sample size is based on typical sample sizes reported in the applied literature. It is entirely possible that some of the tests or models would perform differently if I considered larger sample sizes.

Another limitation in Chapter 2 and 3 is that a specific type of structural break is considered. Specifically, I am interested in the testing and modelling of structural breaks in the mean of the DGP. In the overcharge literature it is often assumed that the cartel will artificially raise prices on its own and not influence the relationship between the determinants of price. This clearly need not always be the case. It is, for example, possible for cartels to influence the cost of inputs through collective bargaining. This will cause a break in the coefficients of variables associated with cost determinants of price. In Chapter 2 some of the results will change if other forms of structural breaks in addition to changes in the mean are considered. Some preliminary simulation evidence indicates that the CUSUM, MOSUM, and sequential F-test perform much better when there is a break in coefficients other than just the mean. In Chapter 3, the differences between the true and t-distribution of the dummy variable coefficient, as well as the size distortions will change. Furthermore, it is not clear how misdating the break dates in this case would influence the cointegration tests and whether similar results will be found.

Practical case data in overcharge estimations is difficult to obtain. In most cases, the data is declared confidential and not publicly available. While I illustrate in Chapter 2 how the simulation findings translate to a practical case, it is still only a single example. It is possible that other cases will find different results due to the nature of the properties in the data. Indeed, in the Sodium chlorate cartel I find that the Bai-Perron tests provide the most reasonable overcharge estimates. However, in Boshoff and Van Jaarsveld (2019) I found that the Bai-Perron tests do not provide plausible overcharge estimates, and instead favour the Markov-switching model.

Practitioners will face a difficult problem when implementing the suggestions, and cautioning involved parties about the findings in this dissertation. Overcharge estimates and methodological approaches followed need to be defendable in court. This requires explanation of complex econometric issues that is not general knowledge for those who do not possess formal training or experience in the subject. As noted in European Commission (2005) and

European Commission (2008) the methodology needs to be kept simple as far as possible. However, as illustrated in this dissertation, adequately dealing with the dating of structural breaks and unit roots can be a complicated undertaking. Practitioners will therefore face the difficult task of assessing the trade-off between complex and simplistic approaches.

A limitation specific to Chapter 2 is not relying on bootstrapped confidence intervals. It is well known that bootstrapping can greatly improve test performance in small samples (Freedman *et al.*, 1981). I am more interested in how these tests and models perform when used as is. One of the general conclusions is that no single measure performs best on every type of structural break. This finding might change if all the tests were bootstrapped.

Note that in Chapter 2, I limit the number of tests and models investigated. Some tests that are not investigated include the use of band spectral regressions, quantile regressions, Lasso-type methods, and continuous-record asymptotics (for an overview see, Casini and Perron, 2019). Again, the choice in the tests and models to investigate, is based on those that were recommended and performed in the applied cartel literature. It might be that a structural break test or regime-switching model exists that will always provide robust estimates regardless of the type of structural break. I therefore do not conclude that no test or model will always provide the best estimate of overcharge, but instead that the tests and models that I investigated do not have a clear winner.

A constraint that is specific to Chapter 3 is that I cannot make use of a case study. To investigate the effects of unit roots on dummy variable coefficients, I require precise knowledge of the DGP and properties of the data. As previously stated, in practice this cannot be declared with absolute certainty. Therefore, the translation of these results to practical applications is not clear. Specifically, there is some uncertainty surrounding the number of cartel cases where an increased price effect was estimated based on incorrect inferential assumptions when there truly was no effect.

As previously discussed, there are two main benchmark approaches used to estimate overcharge. The focus in this dissertation was primarily on the dummy variable approach. I do not formally investigate how structural break tests and unit roots influence the estimates forecasting approach. There are, however, similar issues when using the forecasting approach. The forecasting approach is just as reliant on the correct determination of the break dates. This means that structural break tests that result in biased estimates in the dummy variable approach will have the same effect in the forecasting approach. How the presence of unit roots will affect the forecasting approach is more nuanced. In this case, there will be less of an issue in terms of drawing inference on coefficient estimates. There are, however, two problems to take note of. First, related to the presence of unit roots is the possibility of deterministic trends. If the data contains a deterministic trend that changed direction when the cartel influenced prices, the forecasting approach will no longer be appro-

priate (Davis and Garcés, 2009, 359). Second, the forecasting approach only uses data outside of the contravention period. This can result in exceedingly small samples which can have large effects on the distributions of various test statistics, including cointegration tests, used to test if the model is correctly specified.

5.5 Future research

To further advance the field, there are several questions to consider in future research. I plan to – where possible – systematically address each of the limitations discussed in section 5.4. I do not revisit these limitations in this section, but instead, I discuss related unanswered research questions that are important to consider in the future.

In chapter 2 it would be interesting to see how different structural break tests will perform in similar cases. For example, testing for breaks in quantile regressions (Qu, 2008; Oka and Qu, 2011), can provide interesting insights into how the cartel affected prices in different quantiles or across quantiles. Tests that consider band spectral regressions and low frequency changes could also provide interesting insights (Perron and Yamamoto, 2013). These tests can provide insights into whether a price relationship is stable over the business cycle band. This test is well suited to empirically investigate how the business cycle influence collusion which is a contentions issue in the industrial organization literature (Levenstein and Suslow, 2011, 2016).

An alternative modelling strategy that can estimate both the timing of structural breaks as well as the effects are Smooth Transition Autoregressive models. These models are usually thought of in terms of thresholds in variables. However, recasting them in terms of thresholds in time could provide interesting insights. it

A natural extension to Chapter 2 would be to evaluate how structural break tests, that can accommodate unit roots, perform when used to construct dummy variables in subsequent regressions. I have provided extensive arguments on the importance of testing for structural change in overcharge estimations. However, there is no guidance in the overcharge literature on how to approach structural break tests when the data contains unit roots. This would be an important contribution, since unit roots are often present in practical datasets. Understanding how structural break tests, in the context of unit roots, translate into coefficient estimates of overcharge estimation is therefore necessary.

An issue that is closely related to the unit root and structural break literature is processes with long memory (for an introduction to this topic see, Beran, 1994). This property can be defined as a case where dependence between successive observations decays slowly as the number of lags tend to infinity. A simple example is fractionally integrated processes. These are processes that

show neither evidence of being I(0) or I(1), but have significant autocorrelation up to very long lags. A symptomatic feature of long memory is that it can be caused by structural changes. Long memory and structural breaks have similar data characteristics, since both cases will experience level shifts that would induce persistent features in the data.

If a process has a long memory, there are similar asymptotic distribution limitations than when a unit root is present. It would be interesting to extend the evaluation in Chapter 3 to see what the distribution distortion of a dummy variable coefficient is in the case of a long-memory process. Comparing these results to the unit root case could provide insight into whether the distributional differences are influenced by the strength of integration.

In the context of overcharge, combining the implications of long memory with lingering effects would provide useful insights. Harrington Jr (2004) discusses the issue of prices returning slowly to competitive levels in the post-cartel period. This is correlated with the level of industry concentration and length of the cartel. Testing if various cartels caused the price series to display long memory properties, would have substantial policy implications. If a cartel caused a price series to contain a long-memory the implication is that the effects of the cartel will take an extremely long time to dissipate. This can open the debate to what extent cartels should be held liable for damages caused outside of the formal collusion period.

Appendices

Appendix A

Robustness of cartel dating Monte Carlo simulations

In this appendix we perform additional Monte Carlo (MC) simulations to test the robustness of the results in section 2.5. The data is simulated in a similar fashion than outlined in section 2.4.1. We again apply the Bai-Perron break test (BP), cumulative sum (CUSUM), moving sums (MOSUM), F-tests, Markov-switching (MS) and time-varying parameter (TVP) models to the simulated data. Similar to the results reported in section 2.5, we again consider the four cases of DGPs, namely, deterministic shift, probabilistic shift, recurrent shifts and a shift with transition phases.

In this appendix the results differs from that of section 2.5 where we now specify different parameter values. There are three different combinations that we consider. First, we consider the case where the mean shift is small. Second we consider a case where there is stronger persistence in the time-series than what was considered in chapter 2. Last, we consider the case where the change in intercept is small and the persistence is large.

For brevity, this section does not report the same summary statics tables as was presented in section 2.5.1 to 2.5.4. Instead, the results are only summarized in terms of the bias and relative efficiency.

A.1 Small mean shift

In this case, the mean shift is much smaller than in section 2.5.1 where the parameter of the dummy variable was set equal to 10.

A.1.1 Case 1 - Deterministic shift in DGP

In this setting we generated data using the following DGP:

$$p_t = 50 + 5D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{A.1}$$

with $\epsilon_t \sim IN(0, 1)$.

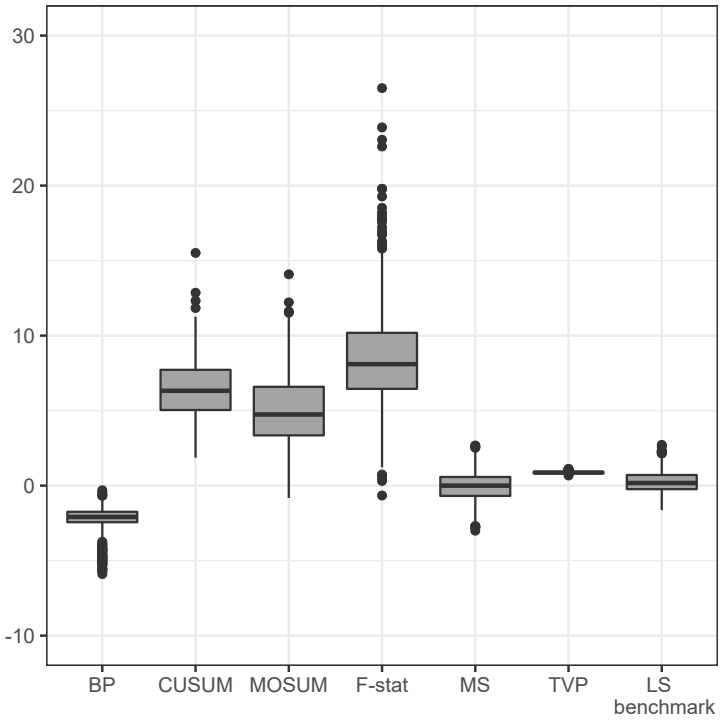


Figure A.1: Bias comparison for deterministic DGP

Table A.1: Relative efficiency for deterministic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
3.14	91.25	60.29	155.32	1.77	1.46

A.1.2 Case 2 - Probabilistic shift in DGP

For the second case we consider the case where the DGP is from an MS model. That is when the DGP is specified as:

$$p_t = \begin{cases} 50 + 0.5x_t + 0.5p_{t-1} + \epsilon_t \text{ for } S_t = 1 \\ 50 + 5 + 0.5x_t + 0.5p_{t-1} + \epsilon_t \text{ for } S_t = 2 \end{cases} \tag{A.2}$$

where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$ and the switches between states evolve according to a first order Markov chain with the following probabilities:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \tag{A.3}$$

A.1.3 Case 3 - Recurrent shifts in the DGP

For this case the DGP has two shifts in the mean where the process enters the collusive period twice. In this case simulate the following DGP

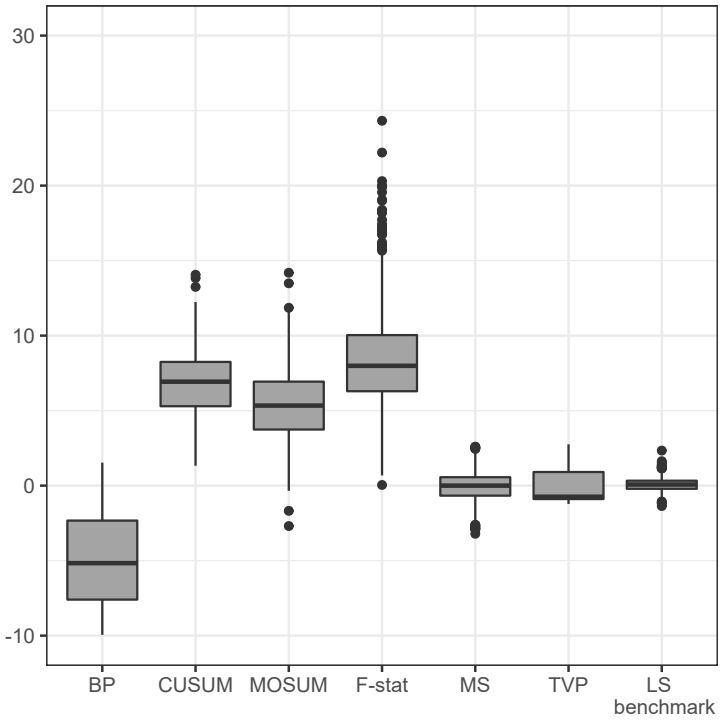


Figure A.2: Bias comparison for probabilistic DGP

Table A.2: Relative efficiency for probabilistic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
205.63	289.21	199.37	452.59	5.28	7.30

$$p_t = 50 + 5D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{A.4}$$

with $\epsilon_t \sim IN(0,1)$ and $D_t = 1$ for $t = 11, ..., 30, 41, ..., 60$ and 0 elsewhere.

Table A.3: Relative efficiency for recurrent DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
37.76	99.76	40.59	124.48	1.75	0.75

A.1.4 Case 4 - Shifts with transition phases

This case represents our DGP that contains transition phases when moving in and out of regimes. For this case the DGP takes the following form

$$p_t = 50 + 5D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t, \tag{A.5}$$

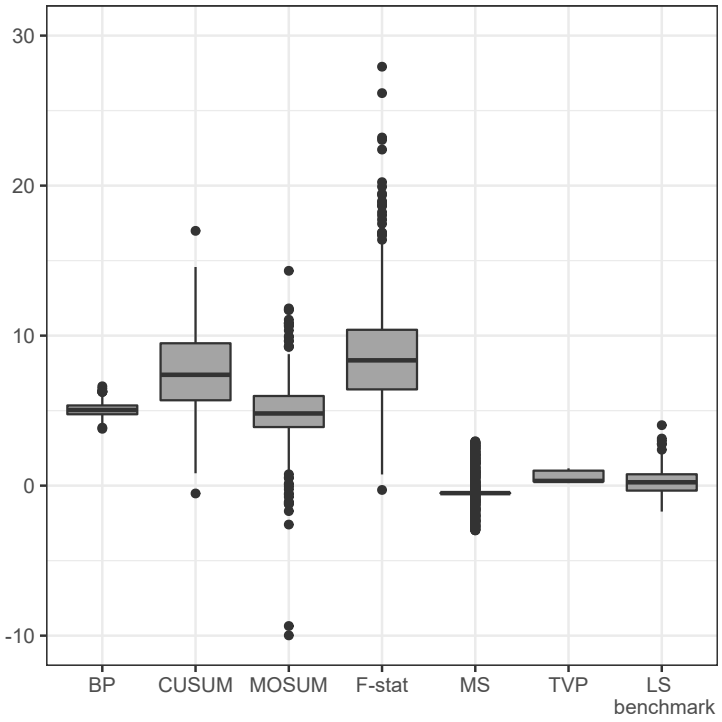


Figure A.3: Bias comparison for recurrent shifts in the DGP

with $\epsilon_t \sim IN(0,1)$ and D_t is encoded as follow:

$$D_t = 0 \text{ for } t = 1, ..., 20, 80, ..., 100$$
$$D_t = 0.1, 0.2, ..., 0.9 \text{ for each } t \text{ in } [21, 22, ..., 29]$$
$$D_t = 1 \text{ for } t = 30, ..., 70$$
$$D_t = 0.9, 0.8, ..., 0.1 \text{ for each } t \text{ in } [71, 72, ..., 79]$$

Table A.4: Relative efficiency for shifts with transition phases

BP	CUSUM	MOSUM	F-stat	MS	TVP
51.14	46.58	32.00	138.77	1.03	0.74

A.2 High degree of persistence

For this section we increase the degree of persistence in the time series.

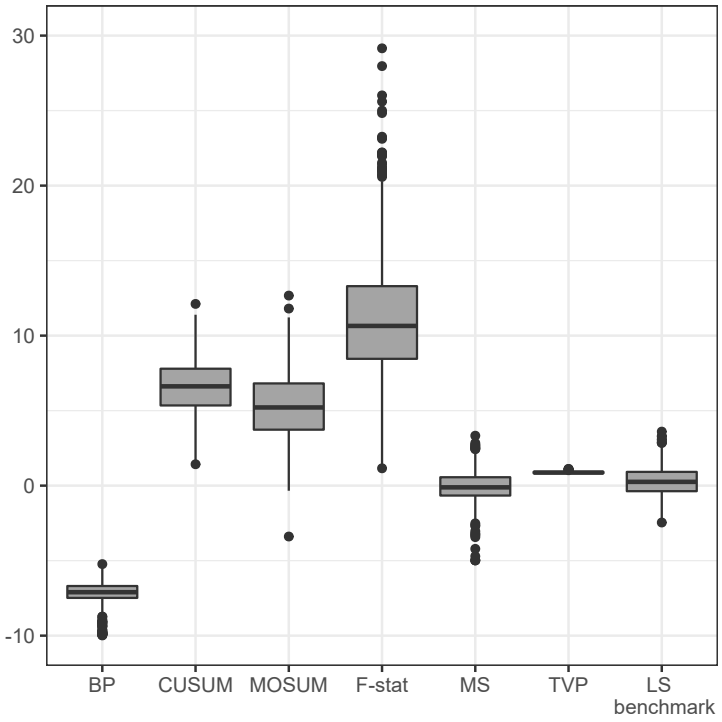


Figure A.4: Bias comparison for shifts with transition phases in the DGP

A.2.1 Case 1 - Deterministic shift in DGP

In this setting we generated data using the following DGP:

$$p_t = 50 + 10D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t \tag{A.6}$$

with $\epsilon_t \sim IN(0, 1)$.

Table A.5: Relative efficiency for deterministic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
0.05	51.02	0.78	0.15	0.04	0.04

A.2.2 Case 2 - Probabilistic shift in DGP

For the second case we consider the case where the DGP is from an MS model. That is when the DGP is specified as:

$$p_t = \begin{cases} 50 + 0.5x_t + 0.8p_{t-1} + \epsilon_t \text{ for } S_t = 1 \\ 50 + 10 + 0.5x_t + 0.8p_{t-1} + \epsilon_t \text{ for } S_t = 2 \end{cases} \tag{A.7}$$

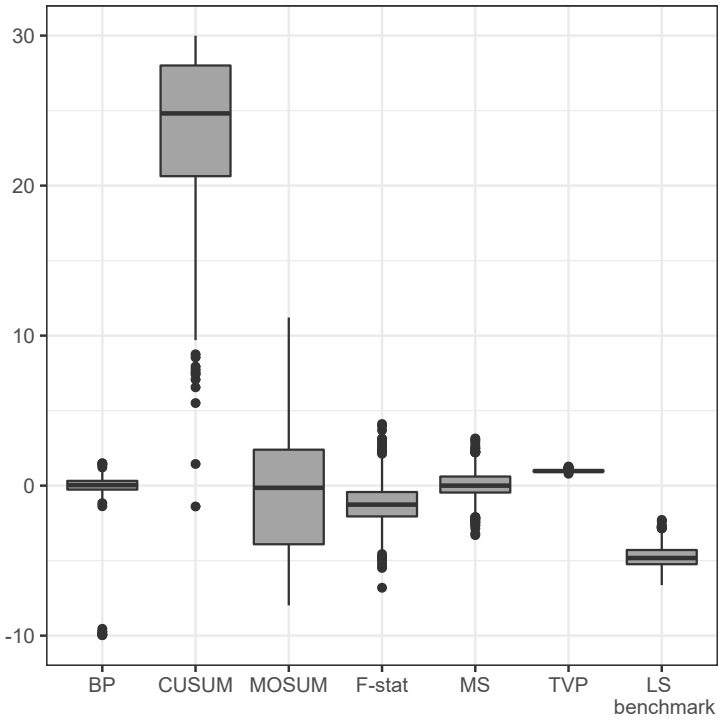


Figure A.5: Bias comparison for deterministic DGP

where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$ and the switches between states evolve according to a first order Markov chain with the following probabilities:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \tag{A.8}$$

Table A.6: Relative efficiency for probabilistic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
486.04	8926.98	127.15	23.94	7.10	8.55

A.2.3 Case 3 - Recurrent shifts in the DGP

For this case the DGP has two shifts in the mean where the process enters the collusive period twice. In this case simulate the following DGP

$$p_t = 50 + 10D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t \tag{A.9}$$

with $\epsilon_t \sim IN(0, 1)$ and $D_t = 1$ for $t = 11, ..., 30, 41, ..., 60$ and 0 elsewhere.

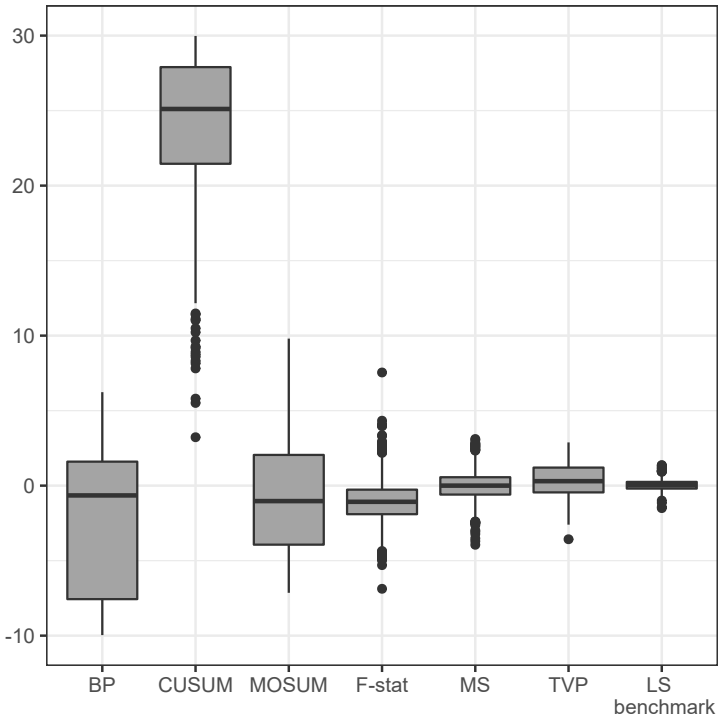


Figure A.6: Bias comparison for probabilistic DGP

Table A.7: Relative efficiency for recurrent DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
0.23	3528.69	6.17	8.06	0.60	5.62

A.2.4 Case 4 - Shifts with transition phases

This case represents our DGP that contains transition phases when moving in and out of regimes. For this case the DGP takes the following form

$$p_t = 50 + 10D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t, \tag{A.10}$$

with $\epsilon_t \sim IN(0, 1)$ and D_t is encoded as follow:

$D_t = 0$ for $t = 1, ..., 20, 80, ..., 100$
$D_t = 0.1, 0.2, ..., 0.9$ for each t in $[21, 22, ..., 29]$
$D_t = 1$ for $t = 30, ..., 70$
$D_t = 0.9, 0.8, ..., 0.1$ for each t in $[71, 72, ..., 79]$

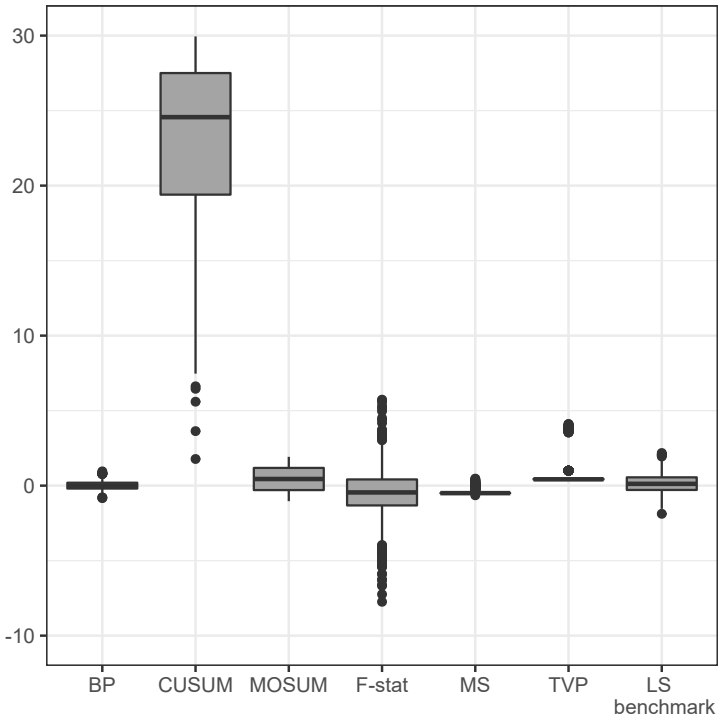


Figure A.7: Bias comparison for recurrent shifts in the DGP

Table A.8: Relative efficiency for shifts with transition phases

BP	CUSUM	MOSUM	F-stat	MS	TVP
163.83	3268.02	48.58	6.66	11.49	5.93

A.3 High degree of persistence and small shifting intercept

For this section we increase the degree of persistence in the time series and have a shifting in.

A.3.1 Case 1 - Deterministic shift in DGP

In this setting we generated data using the following DGP:

$$p_t = 50 + 5D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t \tag{A.11}$$

with $\epsilon_t \sim IN(0, 1)$.

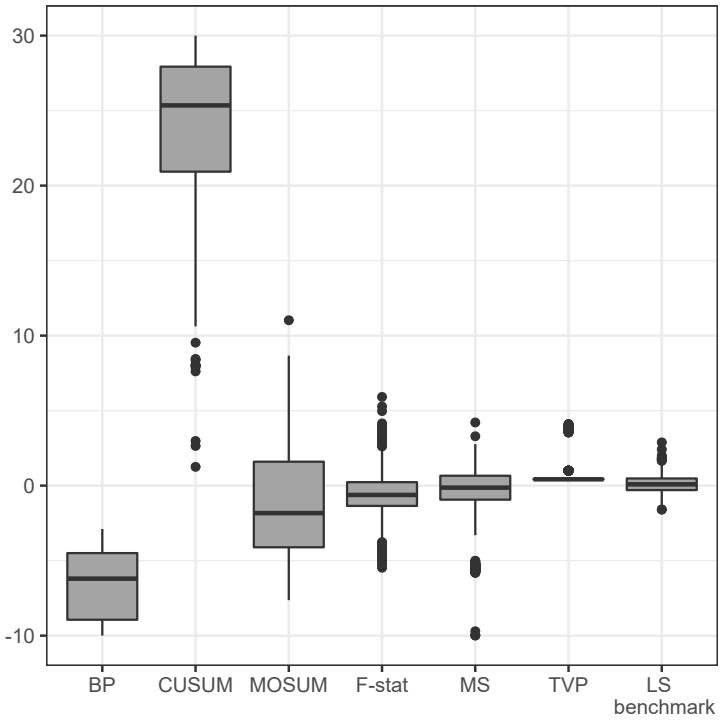


Figure A.8: Bias comparison for shifts with transition phases in the DGP

Table A.9: Relative efficiency for deterministic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
1.33	924.99	37.45	39.19	2.30	0.40

A.3.2 Case 2 - Probabilistic shift in DGP

For the second case we consider the case where the DGP is from an MS model. That is when the DGP is specified as:

$$p_t = \begin{cases} 50 + 0.5x_t + 0.8p_{t-1} + \epsilon_t \text{ for } S_t = 1 \\ 50 + 5 + 0.5x_t + 0.8p_{t-1} + \epsilon_t \text{ for } S_t = 2 \end{cases} \tag{A.12}$$

where $\epsilon_t \sim NI(0, \sigma_\epsilon^2)$ and the switches between states evolve according to a first order Markov chain with the following probabilities:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix} \tag{A.13}$$

A.3.3 Case 3 - Recurrent shifts in the DGP

For this case the DGP has two shifts in the mean where the process enters the collusive period twice. In this case simulate the following DGP

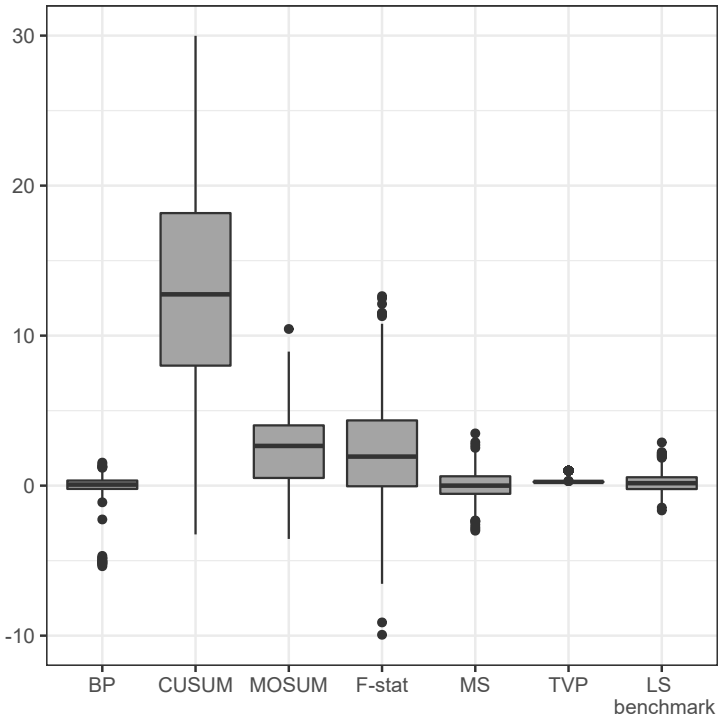


Figure A.9: Bias comparison for deterministic DGP

Table A.10: Relative efficiency for probabilistic DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
283.69	2560.71	101.61	107.62	6.68	28.40

$$p_t = 50 + 5D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t \tag{A.14}$$

with $\epsilon_t \sim IN(0,1)$ and $D_t = 1$ for $t = 11, ..., 30, 41, ..., 60$ and 0 elsewhere.

Table A.11: Relative efficiency for recurrent DGP

BP	CUSUM	MOSUM	F-stat	MS	TVP
0.26	1375.75	6.29	66.72	0.61	1.89

A.3.4 Case 4 - Shifts with transition phases

This case represents our DGP that contains transition phases when moving in and out of regimes. For this case the DGP takes the following form

$$p_t = 50 + 10D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t, \tag{A.15}$$

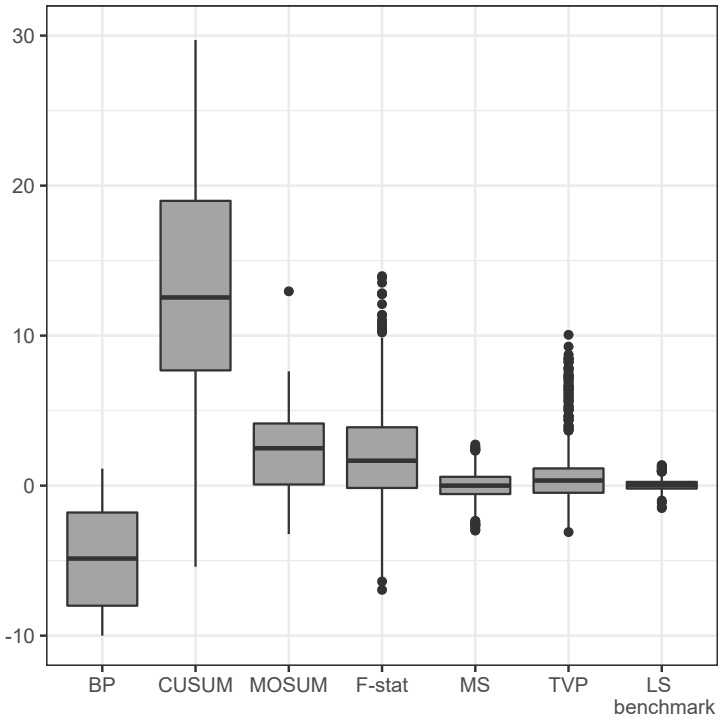


Figure A.10: Bias comparison for probabilistic DGP

with $\epsilon_t \sim IN(0,1)$ and D_t is encoded as follow:

$D_t = 0$ for $t = 1, ..., 20, 80, ..., 100$
$D_t = 0.1, 0.2, ..., 0.9$ for each t in $[21, 22, ..., 29]$
$D_t = 1$ for $t = 30, ..., 70$
$D_t = 0.9, 0.8, ..., 0.1$ for each t in $[71, 72, ..., 79]$

Table A.12: Relative efficiency for shifts with transition phases

BP	CUSUM	MOSUM	F-stat	MS	TVP
16.14	852.87	31.83	72.32	2.60	1.95

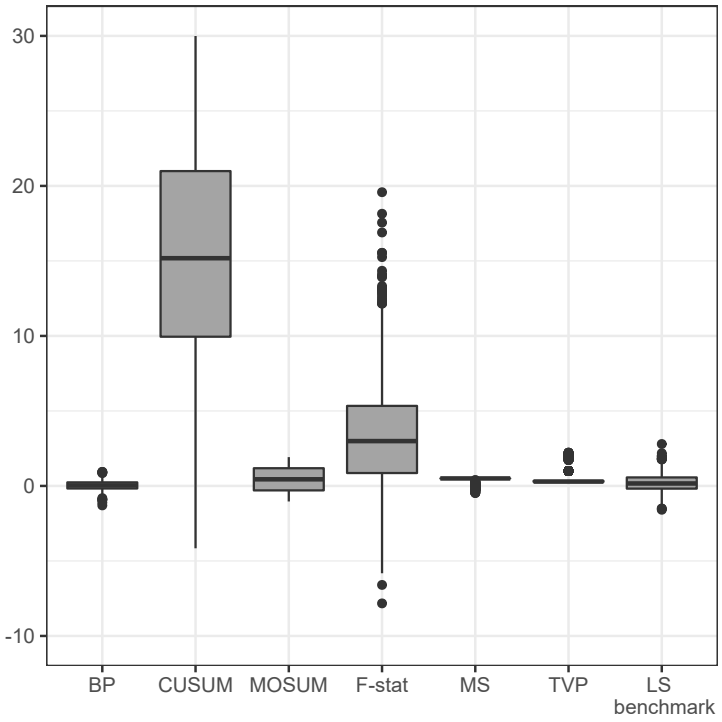


Figure A.11: Bias comparison for recurrent shifts in the DGP

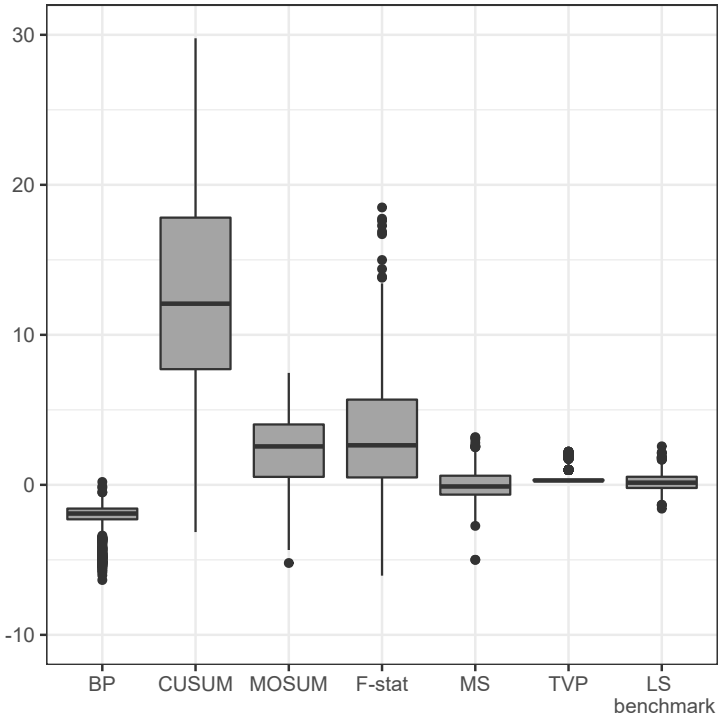


Figure A.12: Bias comparison for shifts with transition phases in the DGP

Appendix B

Residual structure for breaks in the mean

Suppose that the data evolves according to the following DGP

$$p_t = 50 + 10D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{B.1}$$

with $\epsilon_t \sim IN(0,1)$, $T = 100$ and $D_t = 1$ for $t = 25, ..., 50$ and $D_t = 0$ for $t = 1, ...24, 51, ..., 100$. Suppose then that we estimate the following form:

$$p_t = \hat{\alpha}_1 + \hat{\beta}x_1 + \hat{\gamma}p_{t-1} + v_t \tag{B.2}$$

The result from estimating this incorrect form is that the LS estimator will overestimate γ i.e. estimate a γ close to 1. The reason for this occurrence is that it minimizes the residual sum of squares. A graphical example is helpful in this case. We simulate data that has the DGP given in B.1 where $x_t = 0.5x_{t-1} + u_t$. Then, in this single simulated case p_t has the values as shown in Figure B.1.

Now, if we fit a model without a dummy variable, we obtain the following estimates:

	Estimate	Std. Error	t-value
α_1	2.19	2.13	1.03
β	0.71	0.16	4.42
γ	0.98	0.02	50.67

The residuals from the resulting fit is illustrated in Figure B.2.

As can be seen from Figure B.2, the resulting residuals have only two outliers. This occurred due to the high estimated parameter of γ. Because γ is close to 1, the only points at which the fitted values that would be far removed from the actual values are the points at which the breaks occur. The reason for this is because p_{t-1} would incorporate a value that is from a process with a lower mean. The question is, Why does the high γ parameter minimize

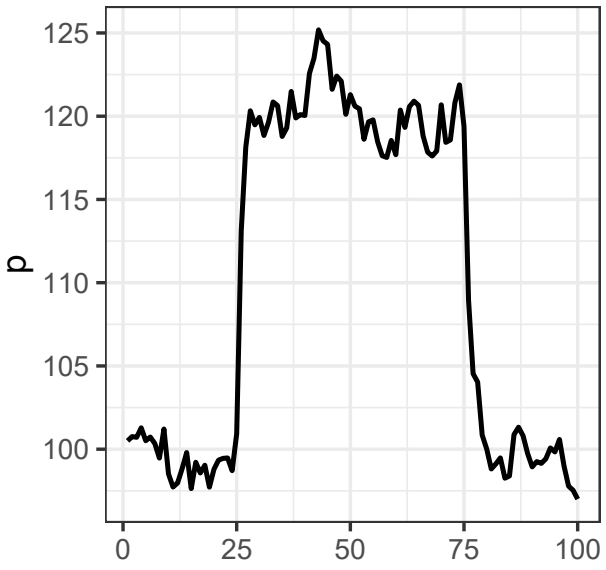


Figure B.1: Example of a single DGP simulation

the sum of squared residuals? To understand why this minimizes the residual sum of square errors, consider what would happen if the parameter values from B.2 are the same as the true parameters from equation B.1. In this case, the residuals will be calculated as:

$$\epsilon_t = y - 50 - 0.5x_t - 0.5y_{t-1} \tag{B.3}$$

What will end up happening is the structural break that is not controlled for during the time period $t = 25, ..., 50$ will raise the residuals by exactly the same amount as the structural break during the period $t = 25, ..., 50$. This is illustrated in Figure B.3.

As can be seen in Figure B.3, during the period $t = 25, ..., 50$ the mean of the residuals increased by 10, which is the exact level shift in the intercept of the true DGP. It is simple to see that the residual sum of squares will be smaller when γ was higher than when the parameter values are closer to the true values.

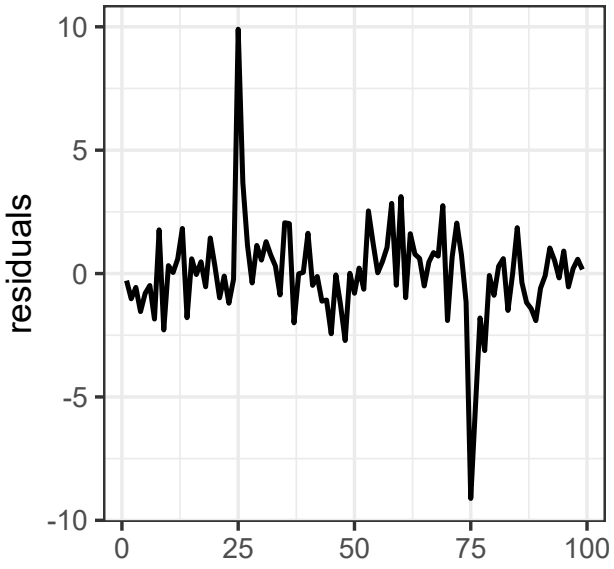


Figure B.2: Example residuals from LS fit

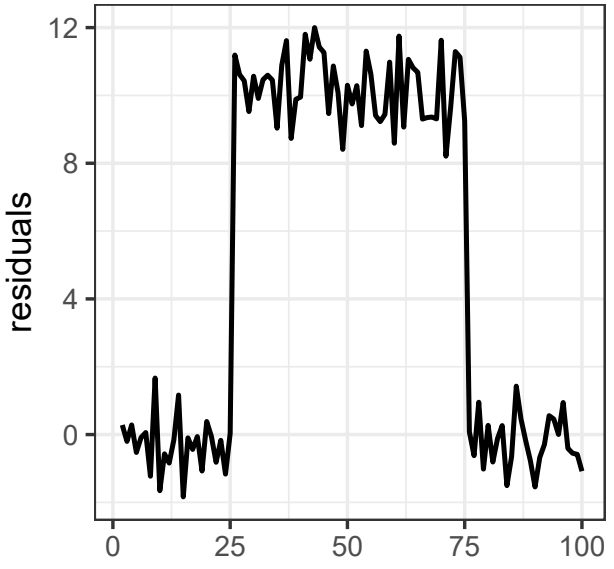


Figure B.3: Example of residuals constructed according to equation B.3

Appendix C

Monte Carlo Markov-switching simulations with estimated changes in autoregressive parameters

This section illustrates two interesting results that we discovered when running the MC experiments on the MS model. To the best of our knowledge these results are not explicitly explained in the literature and require further investigation.

The first result is the case where we simulate a DGP that contain a deterministic shift in the intercept. Additionaly, we estimate a corresponding MS model where we allow for both the intercept and autoregressive parameter to change regimes. The simulated DGP is of the following form:

$$p_t = 50 + 10D_t + 0.5x_1 + 0.5p_{t-1} + \epsilon_t \tag{C.1}$$

with $\epsilon_t \sim IN(0,1)$, $T = 100$ and $D_t = 1$ for $t = 25, ..., 50$ and $D_t = 0$ for $t = 1, ...24, 51, ..., 100$. We then estimate the following MS model to the data from equation C.1:

$$p_t = \alpha_{S_t} + \beta x_1 + \gamma_{S_t} p_{t-1} + \epsilon_t \tag{C.2}$$

where α_{S_t} and γ_{S_t} are the regime dependent intercepts and autoregressive parameters. Figure C.1 illustrates the distributions of the parameters from the simulation.

From Figure C.1, it is clear that the estimated parameters are close to the true DGP specification. The autoregressive parameter, which in the true DGP was not regime dependent, displays interesting behaviour in this case. The estimated values are around the true parameter value of 0.5 with slight underestimation in one regime and slight overestimation in another.

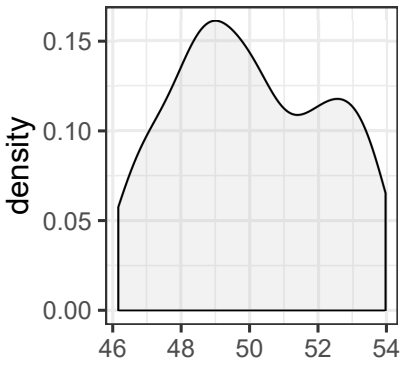
The second interesting finding is when the DGP, from which the data is simulated, has a higher autoregressive parameter value. To illustrate this case, we simulate data from the following DGP:

$$p_t = 50 + 10D_t + 0.5x_1 + 0.8p_{t-1} + \epsilon_t \tag{C.3}$$

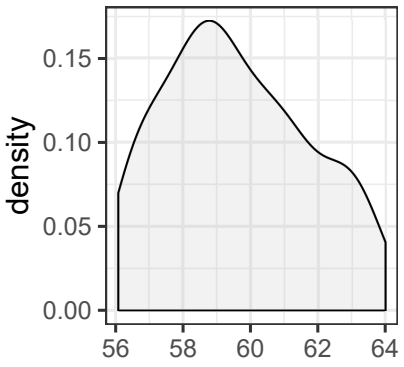
We subsequently fit a MS model with changes in the intercept. Figure C.2 reports the resulting parameter distributions from the simulation.

From Figure C.2 the MS model parameters consistently estimate values that are close to the true parameter values.

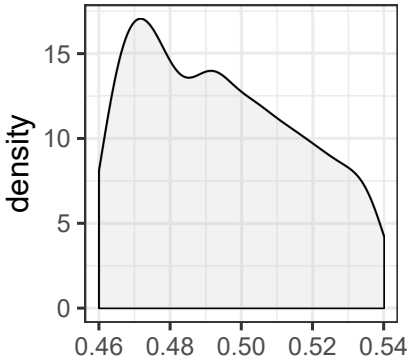
Both of the results presented in this appendix raises important questions around MC simulations in the context of MS simulations. The results show that depending on the MS model specification and chosen parameter values, the ease of estimation in MC simulations changes significantly. That is, for certain combinations of parameter values or MS specification, the optimization complications surrounding the estimation of MS models disappear.



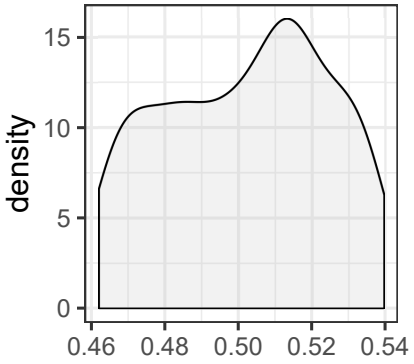
(a) α in regime 1



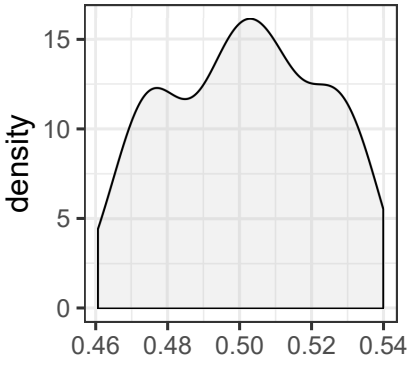
(b) α in regime 2



(c) β

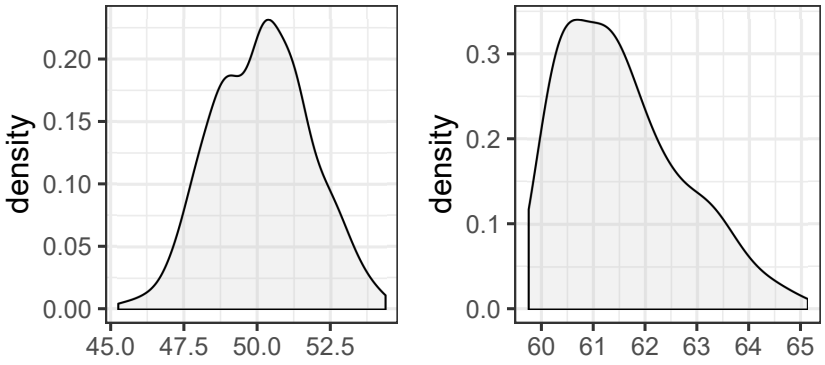


(d) γ in regime 1



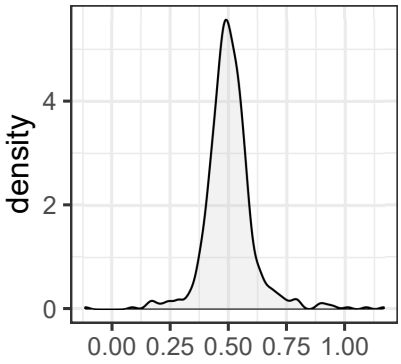
(e) γ in regime 2

Figure C.1: Parameter distributions for simulation when γ is regime dependent

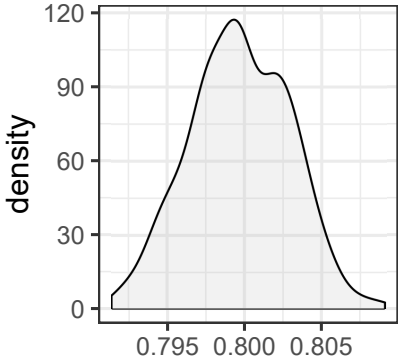


(a) α in regime 1

(b) α in regime 2



(c) β



(d) γ in regime 1

Figure C.2: Parameter distributions for simulation when $\gamma = 0.8$

Appendix D

Effects of nonstationarity on the forecasting approach

Similar to Chapter 2, consider a simple DGP of prices:

$$p_t = \alpha_1 + \alpha_2 D_t + \beta x_t + \gamma p_{t-1} + \epsilon_t \tag{D.1}$$

where p_t is the product unit price in period t, D_t the cartel dummy variable, x_t is a set of explanatory variables and ϵ_t an error term. The sample period is $\mathbb{T} = \{1, ..., T\}$, which can be divided into periods with and without cartel effects, labelled $\mathbb{T}_C$ and $\mathbb{T}_N$ respectively, with $\mathbb{T} = \mathbb{T}_C \cup \mathbb{T}_N$.

In the forecasting approach, only data consisting of $\mathbb{T}_N$ will be used to estimate the parameter values in the following equation:

$$p_t = \alpha_1 + \beta x_t + \gamma p_{t-1} + \epsilon_t \tag{D.2}$$

When dealing with unit roots, trimming the sample in this manner can be problematic for several reasons. Despite the super-consistency of the least-squares estimator of the cointegrating vector, the estimates can be biased in small samples Banerjee *et al.* (1986). By removing all information contained in the sample period $\mathbb{T}_C$ the bias can be further exasperated. Also, it should be noted that as the sample gets smaller, there is a significant power loss in single equation cointegration tests Banerjee (1993). As a result, it can become difficult to identify if a cointegrating relationship exists. Another issue is how this approach performs when deterministic time-trends are involved. If the trend changed direction during the period $\mathbb{T}_C$ the forecast will extrapolate the trend that was present in the period $\mathbb{T}_N$ which will be moving in the incorrect direction. Figure D.1 illustrates this scenario.

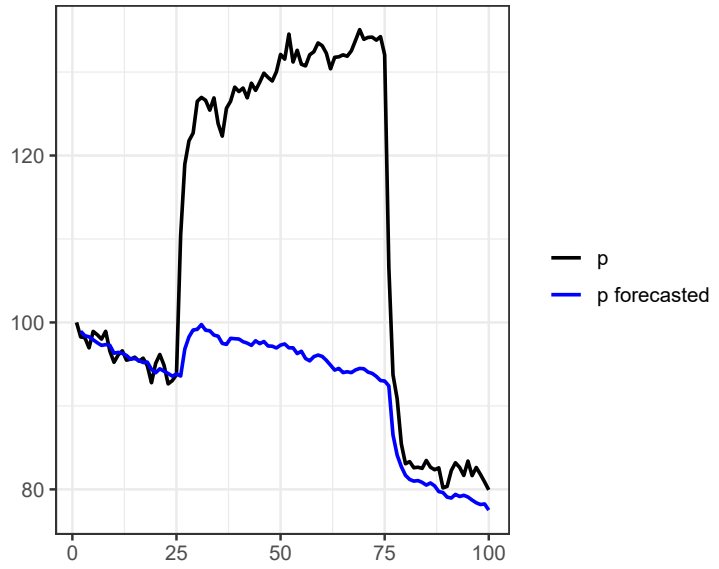


Figure D.1: Forecasting approach with changing trends

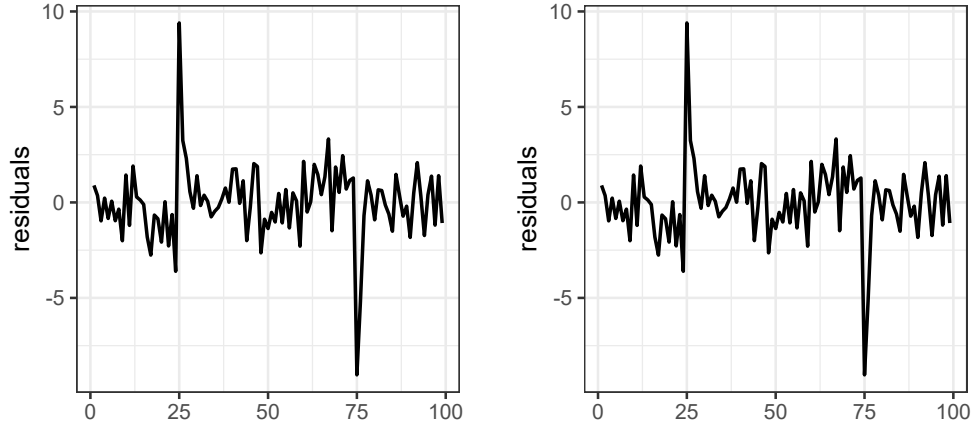
The deterministic trend problem illustrated in Figure D.1 is illustrative of a general shortcoming of the forecasting approach. That is, if any of the parameters of D.1 changed over the period $\mathbb{T}_C$, it is methodologically impossible to control for the change. In other words, if the behaviour of the cartel altered the relationship between price and any of its determinants, the forecasting approach cannot control for this change and the results will be invalid.

From section 3.3, we know that the asymptotic distributions of the parameters in equation D.2 will not be asymptotically normal. This complicates the assessment of the validity of the specification, since standard inference using t-statistics do not apply.

Appendix E

Residual-based cointegration tests in the presence of misspecified dummy variables

To shed some light on why the residual-based cointegration test has a 100% rejection rate of the null hypotheses, Figure E.1 presents the residuals from two regressions where the misspecification was at the start and end of the sample.



(a) Misdating at start of sample (b) Misdating at end of sample

Figure E.1: Residuals from single fit with misdated dummy

As can be seen from Figure E.1 when the dummy variable is misspecified the resulting residuals will have two outliers at the start and end of the "true" observations at which the mean shifted. Since the residual-based cointegration tests are merely a test of whether the residuals contain a unit root, the two outliers created by the misspecification are not enough to lead to an acceptance of the null hypothesis. The reason that there are only two outliers is due to

the fact that OLS would estimate a high autoregressive parameter (close to 1) for the dependent variable[1], since the break is not adequately controlled for. The high autoregressive parameter on the dependent variable will cause the two outliers in the residual since there will be two points where the previous observation belonged to a different regime. Due to this form of behaviour of OLS there will always only be two outliers in the residuals when the mean shift is misspecified.

[1]See full discussion in appendix B

Appendix F

Diagnostic tests for models in the practical application

In chapter 4 various structural break tests are conducted and used to construct a dummy variable that is used in a subsequent regression. This appendix provides diagnostic tests for each of the regression models that is used to estimate overcharge in chapter 4.

The null hypothesis is that there is no serial correlation of any order up to p e null hypothesis that the errors are serially uncorrelated nder the null hypothesis of homoskedasticity null hypothesis that the variances of the errors are not constant

Table F.1: Diagnostic tests for Bai-Perron informed model

test	statistic	p-value
Breusch-Godfrey	14.3486	0.279
Durbin-Watson	1.9074	0.1582
Breusch-Pagan	0.7807	0.3769
Goldfeld-Quandt	1.0331	0.4554

Table F.2: Diagnostic tests for CUSUM informed model

test	statistic	p-value
Breusch-Godfrey	10.2339	0.5954
Durbin-Watson	1.9953	0.3477
Breusch-Pagan	6.9043	0.3298
Goldfeld-Quandt	0.8796	0.7047

A challenge that is faced when performing diagnostic tests for MS models is that the true residuals are unobserved, as they are dependent on the unobserved state variable. To overcome this issue we follow the methodology

Table F.3: Diagnostic tests for MOSUM informed model

test	statistic	p-value
Breusch-Godfrey	8.6229	0.7347
Durbin-Watson	2.06754	0.497
Breusch-Pagan	10.4014	0.1669
Goldfeld-Quandt	0.9953	0.508

Table F.4: Diagnostic tests for sequential F-test informed model

test	statistic	p-value
Breusch-Godfrey	10.1359	0.604
Durbin-Watson	1.997	0.3286
Breusch-Pagan	10.3708	0.1685
Goldfeld-Quandt	0.8467	0.7559

outlined in Maheu and McCurdy (2000) where expected residual are calculated, conditioned on past information.

Table F.5: Diagnostic tests for MS model

test	statistic	p-value
Breusch-Godfrey	9.1989	0.6858
Durbin-Watson	2.001	0.333
Breusch-Pagan	2.6682	0.1024
Goldfeld-Quandt	1.2008	0.2229

Table F.6: Diagnostic tests for TVP model

	test	statistic	p-value
1	Breusch-Godfrey	0.0075	0.9314
2	Durbin-Watson	1.9644	0.3908
3	Breusch-Pagan	0.0197	0.8883
4	Goldfeld-Quandt	0.1845	0.9999

List of References

Andrews, D.W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, vol. 61, no. 4, pp. 821–856.

Andrews, D.W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pp. 1383–1414.

Antoshin, S., Berg, A. and Souto, M. (2008). *Testing for Structural Breaks in Small Samples*. 8-75. International Monetary Fund.

Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory*, vol. 11, no. 3, pp. 403–436.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, vol. 66, no. 1, pp. 47–78.

Bai, J. and Perron, P. (2003*a*). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, vol. 18, no. 1, pp. 1–22.

Bai, J. and Perron, P. (2003*b*). Critical values for multiple structural change tests. *The Econometrics Journal*, vol. 6, no. 1, pp. 72–78.

Banerjee, A. (ed.) (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Advanced texts in econometrics. Oxford University Press, New York.

Banerjee, A., Dolado, J.J., Hendry, D. and Smith, G. (1986). Exploring equilibrium relationships in econometrics through static models: some monte-carlo evidence. *Oxford Bulletin of Economics and Statistics*, vol. 48, no. 3, pp. 207–223.

Bataa, E., Osborn, D.R., Sensier, M. and van Dijk, D. (2013). Structural breaks in the international dynamics of inflation. *The Review of Economics and Statistics*, vol. 95, no. 2, pp. 646–659.

Beck, N. (1983). Time-varying parameter regression models. *American Journal of Political Science*, vol. 27, no. 3, pp. 557–600.

Beran, J. (1994). *Statistics for Long-Memory Processes*, vol. 61. CRC press.

Boshoff, W.H. (2015). Illegal cartel overcharges in markets with a legal cartel history: Bitumen prices in s outh a frica. *South African Journal of Economics*, vol. 83, no. 2, pp. 220–239.

Boshoff, W.H. and van Jaarsveld, R. (2018). Recurrent collusion: Cartel episodes and overcharges in the south african cement market. *Review of Industrial Organization*, vol. 54, no. 2, pp. 353–380.

Boshoff, W.H. and Van Jaarsveld, R. (2019). Recurrent collusion: Cartel episodes and overcharges in the south african cement market. *Review of Industrial Organization*, vol. 54, no. 2, pp. 353–380.

Boswijk, H.P., Bun, M.J. and Schinkel, M.P. (2019). Cartel Dating. *Journal of Applied Econometrics*, vol. 34, no. 1, pp. 26–42.

Breiman, L. *et al.* (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, vol. 16, no. 3, pp. 199–231.

Breunig, R., Najarian, S. and Pagan, A. (2003). Specification testing of markov switching models. *Oxford Bulletin of Economics and Statistics*, vol. 65, pp. 703–725.

Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 37, no. 2, pp. 149–163.

Campos, J., Ericsson, N.R. and Hendry, D.F. (1996). Cointegration tests in the presence of structural breaks. *Journal of Econometrics*, vol. 70, no. 1, pp. 187–220.

Cappuccio, N. and Lubian, D. (2001). Estimation and inference on long-run equilibria: A simulation study. *Econometric Reviews*, vol. 20, no. 1, pp. 61–84.

Casini, A. and Perron, P. (2019). Structural breaks in time series. *Oxford Research Encyclopedia of Economics and Finance*.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, vol. 28, no. 3, pp. 591–605.

Chu, C.-S.J., Hornik, K. and Kuan, C.-M. (1995). The moving-estimates test for parameter stability. *Econometric Theory*, vol. 11, no. 4, pp. 699–720.

Ciuperca, G. (2014). Model selection by lasso methods in a change-point model. *Statistical Papers*, vol. 55, no. 2, pp. 349–374.

Connor, J.M. (2007). Forensic economics: An introduction with special emphasis on price fixing. *Journal of Competition Law and Economics*, vol. 4, no. 1, pp. 31–59.

Connor, J.M. (2014). Price-fixing overcharges: revised 3rd edition. *The Law and Economics of Class Actions: Research in Law and Economics*, pp. 59–153.

Crede, C.J. (2019). A structural break cartel screen for dating and detecting collusion. *Review of Industrial Organization*, vol. 54, no. 3, pp. 543–574.

Davidson, A.C. and Hinkley, D. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.

Davis, P. and Garcés, E. (2009). *Quantitative Techniques for Competition and Antitrust Analysis*. Princeton University Press.

Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, pp. 1057–1072.

Dunn, W.L. and Shultis, J.K. (2012). *Exploring Monte Carlo methods*. Elsevier.

Enders, W. (2008). *Applied Econometric Time Series*. John Wiley & Sons.

Engle, R.F. and Granger, C.W.J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, vol. 55, no. 2, pp. 251–276.

Engle, R.F. and Yoo, B.S. (1987). Forecasting and testing in co-integrated systems. *Journal of econometrics*, vol. 35, no. 1, pp. 143–159.

European Commission (2005). Green paper-damages actions for breach of the ec antitrust rules. *COM(2005) 672*.

European Commission (2008). White paper on damages actions for breach of the ec antitrust rules. *COM(2008) 165*.

European Commission (2013). Quatifying harm in actions for damages based on breaches of Article 101 or 102 of the treaty on the functioning of the European Union. *SWD(2013) 205*.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, vol. 66. CRC Press.

Frank, N. and Lademann, R.P. (2010). Economic evidence in private damage claims: What lessons can be learned from the german cement cartel case? *Journal of European Competition Law & Practice*, vol. 1, no. 4, pp. 360–366.

Freedman, D.A. *et al.* (1981). Bootstrapping regression models. *The Annals of Statistics*, vol. 9, no. 6, pp. 1218–1228.

Friederiszick, H.W. and Röller, L.-H. (2010). Quantification of harm in damages actions for antitrust infringements: Insights from german cartel cases. *Journal of Competition Law and Economics*, vol. 6, no. 3, pp. 595–618.

Gardner, L. (1969). On detecting changes in the mean of normal variates. *The Annals of Mathematical Statistics*, vol. 40, no. 1, pp. 116–126.

Gentle, J.E. (2003). *Random Number Generation and Monte Carlo Methods*. Springer, New York.

Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, pp. 424–438.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, vol. 16, no. 1, pp. 121–130.

Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, vol. 2, no. 2, pp. 111–120.

Green, E.J. and Porter, R.H. (1984). Noncooperative collusion under imperfect price information. *Econometrica: Journal of the Econometric Society*, vol. 52, no. 1, pp. 87–100.

Gregory, A.W. and Hansen, B.E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics*, vol. 70, no. 1, pp. 99–126.

Gregory, A.W., Nason, J.M. and Watt, D.G. (1996). Testing for structural breaks in cointegrated relationships. *Journal of Econometrics*, vol. 71, no. 1-2, pp. 321–341.

Groen, J.J., Kapetanios, G. and Price, S. (2013). Multivariate methods for monitoring structural change. *Journal of Applied Econometrics*, vol. 28, no. 2, pp. 250–274.

Haltiwanger, J. and Harrington Jr, J.E. (1991). The impact of cyclical demand movements on collusive behavior. *The RAND Journal of Economics*, vol. 22, no. 1, pp. 89–106.

Hamilton, J. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, vol. 57, no. 2, pp. 357–384.

Hamilton, J.D. (2010). Regime switching models. In: *Macroeconometrics and time series analysis*, pp. 202–209. Springer.

Hansen, B.E. (1995). Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econometric Theory*, vol. 11, no. 5, pp. 1148–1171.

Hanson, B.E. (1992). Tests for parameter instability in regressions with i (1) processes. *Journal of Business & Economic Statistics*, vol. 10, no. 1, pp. 321–325.

Harrington Jr, J.E. (2004). Post-cartel pricing during litigation. *The Journal of Industrial Economics*, vol. 52, no. 4, pp. 517–533.

Hüschelrath, K., Müller, K. and Veith, T. (2016). Estimating damages from price-fixing: the value of transaction data. *European Journal of Law and Economics*, vol. 41, no. 3, pp. 509–535.

Hüschelrath, K., Müller, K. and Veith, T. (2016). Estimating damages from price-fixing: the value of transaction data. *European Journal of Law and Economics*, vol. 41, no. 3, pp. 509–535.

Jochmann, M. and Koop, G. (2015). Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics*, vol. 19, no. 1, pp. 35–48.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, vol. 12, no. 2-3, pp. 231–254.

Juselius, K. (2017). Recent developments in cointegration. *Econometrics*, vol. 6, no. 1.

Khumalo, J., Mashiane, J. and Roberts, S. (2014). Harm and overcharge in the south african precast concrete products cartel. *Journal of Competition Law and Economics*, vol. 10, no. 3, pp. 621–646.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, vol. 60, no. 1-2, pp. 1–22.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, vol. 4, pp. 83–91.

Krämer, W. (2012). *Econometrics of Structural Change*. Springer Science & Business Media.

Krolzig, H.-M. (2013). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, vol. 454. Springer Science & Business Media.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, vol. 54, no. 1-3, pp. 159–178.

Lee, J., Huang, C.J. and Shin, Y. (1997). On stationary tests in the presence of structural breaks. *Economics Letters*, vol. 55, no. 2, pp. 165–172.

Lee, J. and Strazicich, M.C. (2003). Minimum Lagrange Multiple Unit Root Test With Two Structural Breaks. *The Review of Economics and Statistics*, vol. 84, no. 4, pp. 1082–1089.

Lee, L.-F. and Porter, R.H. (1984 March). Switching Regression Models with Imperfect Sample Separation Information–With an Application on Cartel Stability. *Econometrica*, vol. 52, no. 2, pp. 391–418.

Levenstein, M.C. and Suslow, V.Y. (2011). Breaking up is hard to do: Determinants of cartel duration. *The Journal of Law and Economics*, vol. 54, no. 2, pp. 455–492.

Levenstein, M.C. and Suslow, V.Y. (2016). Price fixing hits home: An empirical study of us price-fixing conspiracies. *Review of Industrial Organization*, vol. 48, no. 4, pp. 361–379.

MacKinnon, J.G. (1991). Critical values for cointegration tests. In: *Eds.), Long-Run Economic Relationship: Readings in Cointegration*. Citeseer.

Maheu, J.M. and McCurdy, T.H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, vol. 18, no. 1, pp. 100–112.

Maki, D. (2011). Pitfalls in estimating cointegrating vector when cointegration relationship has nonlinear adjustment. *Communications in Statistics - Simulation and Computation*, vol. 40, no. 8, pp. 1111–1121.

McConnell, M.M. and Perez-Quiros, G. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *The American Economic Review*, vol. 90, no. 5, pp. 1464–1476.

McCrary, J. and Rubinfeld, D.L. (2014). Measuring Benchmark Damages in Antitrust Litigation. *Journal of Econometric Methods*, vol. 3, no. 1.

Mncube, L. (2014). The south african wheat flour cartel: Overcharges at the mill. *Journal of industry, Competition and Trade*, vol. 14, no. 4, pp. 487–509.

Nelson, C.R. and Plosser, C.I. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of Monetary Economics*, vol. 10, no. 2, pp. 139–162.

Nielsen, H.B. (2004). Cointegration analysis in the presence of outliers. *The Econometrics Journal*, vol. 7, no. 1, pp. 249–271.

Oka, T. and Perron, P. (2018). Testing for common breaks in a multiple equations system. *Journal of Econometrics*, vol. 204, no. 1, pp. 66–85.

Oka, T. and Qu, Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics*, vol. 162, no. 2, pp. 248–267.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, vol. 42, no. 3/4, pp. 523–527.

Page, E. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, vol. 44, no. 1/2, pp. 248–252.

Park, J.Y. and Phillips, P.C. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, vol. 5, no. 1, pp. 95–131.

Park, J.Y. and Phillips, P.C.B. (1988). Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory*, vol. 4, no. 3, pp. 468–497.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, vol. 57, no. 6, pp. 1361–1401.

Perron, P. (1990). Testing for a Unit Root in a Time Series with a Changing Mean. *Journal of Business & Economic Statistics*, vol. 8, no. 2, pp. 153–162.

Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics*, vol. 80, no. 2, pp. 355–385.

Perron, P. and Yamamoto, Y. (2013). Estimating and testing multiple structural changes in linear models using band spectral regressions. *The Econometrics Journal*, vol. 16, no. 3, pp. 400–429.

Perron, P. *et al.* (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics*, vol. 1, no. 2, pp. 278–352.

Pesaran, M.H. and Shin, Y. (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. *Econometric Society Monographs*, vol. 31, pp. 371–413.

Pesaran, M.H., Shin, Y. and Smith, R.J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, vol. 16, no. 3, pp. 289–326.

Phillips, P. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, vol. 33, no. 3, pp. 311–340.

Phillips, P.C. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, pp. 165–193.

Phillips, P.C.B. and Durlauf, S.N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, vol. 53, no. 4, pp. 473–495.

Phillips, P.C.B. and Hansen, B.E. (1990). Statistical inference in instrumental variables regression with i(1) processes. *The Review of Economic Studies*, vol. 57, no. 1, pp. 99–125.

Ploberger, W. and Krämer, W. (1992). The cusum test with ols residuals. *Econometrica: Journal of the Econometric Society*, vol. 60, no. 2, pp. 271–285.

Politis, D.N. and Romano, J.P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1303–1313.

Porter, R.H. (1983). Optimal cartel trigger price strategies. *Journal of Economic Theory*, vol. 29, no. 2, pp. 313–338.

Primiceri, G.E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, vol. 72, no. 3, pp. 821–852.

Prodan, R. (2008). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business & Economic Statistics*, vol. 26, no. 1, pp. 50–65.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, vol. 146, no. 1, pp. 170–184.

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, vol. 53, no. 284, pp. 873–880.

Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, vol. 55, no. 290, pp. 324–330.

Rosenbrock, H. (1960). An automatic method for finding the greatest or least value of a function. *The Computer Journal*, vol. 3, no. 3, pp. 175–184.

Rotemberg, J.J. and Saloner, G. (1986). A supergame-theoretic model of price wars during booms. *The American Economic Review*, vol. 76, no. 3, pp. 390–407.

Rubinfeld, D.L. (2012). Antitrust damages. In: *Research Handbook on the Economics of Antitrust Law*. Edward Elgar Publishing.

Sanchez-Espigares, J. and Lopez-Moreno, A. (2018). Package 'mswm': Fitting markov switching models, r package version 1.4.

Schmidt, P. and Phillips, P. (1992). Testing forunit root in the presence of deterministic trends. *Oxford Bull. Econ. Stat.*, vol. 54, pp. 257–288.

Schwert, G.W. (2002). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 5–17.

Sensier, M. and Dijk, D.v. (2004). Testing for volatility changes in u.s. macroeconomic time series. *The Review of Economics and Statistics*, vol. 86, no. 3, pp. 833–839.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 279–281.

Stock, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, vol. 55, no. 5, pp. 1035–1056.

Stock, J.H. and Watson, M.W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica*, vol. 61, no. 4, pp. 783–820.

Stock, J.H. and Watson, M.W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, vol. 14, no. 1, pp. 11–30.

Tucci, M.P. (1995). Time-varying parameters: a critical introduction. *Structural Change and Economic Dynamics*, vol. 6, no. 2, pp. 237–260.

Vinod, H. (2004). Ranking mutual funds using unconventional utility theory and stochastic dominance. *Journal of Empirical Finance*, vol. 11, no. 3, pp. 353–377.

Vinod, H.D. (2006). Maximum entropy ensembles for time series inference in economics. *Journal of Asian Economics*, vol. 17, no. 6, pp. 955–978.

Vogt, H., Balej, J., Bennett, J.E., Wintzer, P., Sheikh, S.A., Gallone, P., Vasudevan, S. and Pelin, K. (2000). Chlorine oxides and chlorine oxygen acids. *Ullmann's Encyclopedia of Industrial Chemistry*.

Wang, Z. (2006). The joint determination of the number and the type of structural changes. *Economics Letters*, vol. 93, no. 2, pp. 222–227.

Wooldridge, J.M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Xiao, Z. and Phillips, P.C. (1999). Efficient detrending in cointegrating regression. *Econometric Theory*, vol. 15, no. 4, pp. 519–548.

Yule, G.U. (1926). Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society*, vol. 89, no. 1, pp. 1–63.

Zivot, E. and Andrews, D.W.K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, vol. 10, no. 10, pp. 251–70.